Special Negative Database (SNDB) for Protecting Privacy in Big Data

Tamer Abdel Latif Ali[1]*, Mohamed Helmy Khafagy[2], Mohamed Hassan Farrag[3]
Computer Science Department, College of Computing & Information Technology[1]
Arab Academy for Science, Technology and Maritime Transport, Aswan, Egypt[1]
Computer Science Department, Faculty of Computers & Information, Fayoum University, Fayoum, Egypt[1, 2]
Information Systems Department, Faculty of Computers & Information, Fayoum University, Fayoum, Egypt[3]

***Abstract*—Despite the importance of big data, it faces many challenges. The most important big data challenges are data storage, heterogeneity, inconsistency, timeliness, security, scalability, visualization, fault tolerance, and privacy. This paper concentrates on privacy which is one of the most pressing issues with big data. As mentioned in the Literature Review below there are numerous methods for safeguarding privacy with big data. This paper introduces an efficient technique called Specialized Negative Database (SNDB) for protecting privacy in big data. SNDB is proposed to avoid the drawbacks of all previous techniques. SNDB is based on deceiving bad users and hackers by replacing only sensitive attribute with its complement. Bad user cannot differentiate between the original data and the data after applying this technique.**

Keywords—Big data; big data challenges; privacy violations; privacy-preserving techniques; special negative database; data integrity

I. Introduction

One of the most pressing challenges in big data is data privacy. Patients' data must be kept private since there is a risk of improper use of personal information being exposed when data from multiple sources is combined. In Privacy, every person has the right to select the extent of his or her interaction with the environment, as well as the amount of data that can be accessible by a third party. While it is sufficient to detect information as a "password" in security issues, since security is between two trusted parties, the server provider (SP) may be an adversary in privacy difficulties. We classified likely privacy violations in big data systems into four categories based on a literature review: data breaches, re-identification attacks, information gathering by service providers, and government tracking. The motivation of this manuscript is the importance of preserving privacy for everyone specially when dealing with big data. Also, the drawbacks of previous techniques like time consuming, losing data integrity, increasing size of data, low level of privacy and high complexity are one of the motivation factors for the author to propose a new technique called SNDB that will avoid drawbacks of other techniques. The next section will introduce literature review of previous techniques and their drawbacks. While in the third section, proposed technique and the manuscript contribution will be introduced. In fourth section, the author will introduce datasets used in proposed technique. Fifth section will discuss results and evaluation of the proposed technique when comparing with other techniques. Finally, conclusion and future work will be introduced [1], [2], [3], [4].

II. Litrature Review

A. Privacy Preserving by Slicing

Slicing is a method of dividing a dataset in vertical and horizontal manner. The process of dividing attributes into columns based on their correlations means vertical partitioning. Slicing can handle data with high dimensions according to attribute splitting. However, horizontal partitioning happens when records are combined into various buckets, and values in each column are permuted within each bucket randomly to disrupt the relationship between columns. The links between columns are broken by slicing, but the associations within each column are preserved [5].

B. Privacy in Big Data Generation Phase

1) Access restriction: Advertisement blockers, encryption methods, anti-tracking extensions, anti-virus software and anti-Malware are used to limit the access to sensitive data [6].

2) Falsifying data

- Socketpuppet is a deception-based method of masking an individual's internet identity [6].
- Users can use MaskMe to establish aliases for personal information such as their credit card number or email address [6].

C. Privacy in Big Data Storage Phase

1) Attribute based encryption (ABE); ABE is a cloud storage encryption technique that assures big data privacy. The data owner defines the access policies in ABE, and data is encrypted according to those policies. Users whose features match with the data owner's access requirements can decrypt the encrypted data [6].

2) Identity based encryption (IBE): IBE is used to simplify key management in a certificate-based public key infrastructure (PKI) by employing personal identities as public keys, such as an IP address or an email address, to maintain sender and receiver anonymity [6].

3) Homomorphic encryption: By calculating directly on the encryption of a message, it is possible to obtain the encryption of a function of that message [6].

*ta1284@fayoum.edu.eg

4) Storage path encryption: The huge amount of data is first divided into numerous sequential parts, and each component is then saved on a distinct storage media controlled by several cloud storage providers [6].

5) Usage of hybrid clouds: The inherent qualities of public clouds, such as scalability and processing capacity, are combined with the inherent features of private clouds, such as security, to open up possible research opportunities in the processing and storage of enormous amounts of data [6].

D. Privacy in Big Data Processing Phase using Anonymization Techniques

1) Generalization: In the taxonomy of an attribute, a parent value is used to replace some values. An artist, rather than a singer or actor, might be used to symbolise a job attribute [6].

2) Suppression: A special character (e.g., *") is replaced for some values to declare that the modified value is not exposed in suppression. Value suppression, record suppression and cell suppression are examples of suppression schemes [6].

3) Anatomization: Rather of changing the quasi-identifier or sensitive features, anatomization separates the connection between the two [6].

4) Permutation: By dividing a set of data into groups and rearranging the sensitive values within each group, the connection between the quasi-identifier and the numerically sensitive feature is de-associated in permutation [6].

5) Perturbation: The actual data values are replaced with generated data values in perturbation, resulting in statistical information acquired from modified data that is statistically similar to that computed from the original data [6].

E. Privacy Protection Using Laws and Cyber Security

1) Privacy laws and regulations: Regulations and Laws have helped to protect privacy by limiting government tracking and limiting the reading, analysing, and publishing of users' personal information. Laws can also compel service providers to put in place necessary safeguards to protect data confidentiality and prevent data theft. This can enhance protecting privacy by avoiding privacy violations [7].

2) Cyber security measures to prevent data breaches and cyber attacks [7], [8].

- Honeypots and other espionage devices.
- Firewalls and other preventative measures.
- Malicious behavior is also detected via access logs and alert systems.
- Mechanisms for encrypting data.

F. Foggy Dummies

This approach is utilized in fog computing, and the fundamental idea is to create extremely intelligent dummies to preserve the user's privacy. This technique is used by the researcher to swap requests between fogs before sending them to server provider and then swapping the responses. This will be accomplished by fogs cooperating to exchange data before sending it to the server provider [9].

G. Blind Third Party (BTP)

The essential point is why we must rely on a third party (TP) to keep the user safe from SP. That is, we are transferring the problem from one server to another. This strategy is dependent on fog's role as a middleman between the user and the SP in each location [9].

H. Double Foggy Cache

The primary idea behind this method is to use traditional cooperation to tackle the problem of peer trust. Furthermore, use SP to preserve your privacy. This strategy, in particular, can be seen as a significant advancement in the field. To accomplish this, we propose placing two caches in the Fog that will operate as intermediaries between peers. The first is for questions, while the second is for responses [9].

I. Secured Map Reduce Model (SMR)

As the data passes through the map-reduce phase, this new layer applies the security techniques to each individual piece of data. This security technique should be a simple encryption scheme, so that the complexity of new technique does not interfere with the big data's fundamental functioning. When data is processed using this suggested Secured Map Reduce (SMR) layer of big data, it can also be stored and secured. It begins with the collecting of data from social media, weblogs, and streaming data which is then delivered to Hadoop Distributed File System (HDFS). SMR is a suggested paradigm that adds a privacy layer between HDFS and the Map Reduce Layer (MR). Randomized procedures and perturbation were employed to strengthen the data's privacy [10].

J. Blind Peer Approach

This technique fixes the fundamental flaw in the prior technique, in which blind third party may collude with server provider to infringe on consumers' privacy. The new notion in the BLP strategy is to rely on collaboration with a large number of peers rather than dealing with a single TP. As a result of user's request would be sent to another peer in the same area, then encrypted by SPPK, giving the other peer no choice but to pass the question on to the SP, who would decrypt and resolve it [11].

K. Integrated Blind Parties (IBPs)

By integrating the BTP and BLP, This IBPs strategy raises the level of privacy while removing the disadvantages of the other seven options. When a peer isn't active in the area, the user can only rely on the BLP in this case. Furthermore, in the event of a resource shortage, without encrypting the query, the user might exchange it with another peer. In that circumstance, the peer can perform the BTP strategy rather than the user. This strategy can be used in any of the seven techniques [11].

L. Negative Database Conversion Algorithm

Instead of a single tuple, a negative database conversion technique is utilised to generate a big set of values. The data sets that have been generated are inserted into the database. In contrast to normal database applications, a harmful request in our negative database will be unable to access the database's data. Because of the fabrication of fake sets of data in comparison to the premier data, the term negative is utilized. Both database encryption algorithms and virtual database encryption are used to encrypt the actual data [12].

M. Negative Database and Generic Database

The Entity, Attributes, and Values model of the general database design (EAV) was evaluated using the blob data storage type. The data collected, such as exam results, will be organised into three columns Entity, Attributes and Values as the name implies, the EAV is made up of three parts: entities, attributes, and values. The most straightforward approach to apply this principle is to create three tables for each data input (entity). There are two ways to implement the Negative database concept: one that statically generates negative data, i.e. the System Administrator defines the Negative data. Another that both statically and dynamically generates negative data. The user I.e. generates the dynamic negative data [13].

N. Enhanced BTP

In this enhanced approach, there is a new factor added to the old BTP, which is a unique token. This new technique consists of seven factors. A unique token is defined when the user sends a hidden code within a query to the service provider (SP) while SP returns the previous query token. Then SP will store the token for each ID generated by the third party, so the previous one cannot be used in a later query. When the third party inquiries from SP, a change will occur on the user's token, and the user will discover unauthorized access to his data by third party, so the proposed technique will be a powerful guarantee that there is no breakthrough [14].

O. Light Weight Cryptography Techniques(LWCT)

Based on an oil spill detection application, LWCT is utilized to secure a data transmission framework for the internet of things. Through locative and boundary value aggregation, this strategy eliminates duplicate data transmission. The suggested method protects data transfer by combining known lightweight cryptographic techniques with simple ID-based authentication [15], [16].

P. Block Nested Loop (BNL) Skyline Algorithms

This method is used to determine which encryption algorithm is best for ensuring data protection and privacy issue. The author of the Skyline algorithm considers two primary parameters: the rate of variation and the number of dimensions [17].

The author summaries all important drawbacks of previous techniques in the following factors:

- Time consuming
- High complexity
- Losing data integrity
- Low privacy protection
- Increasing size of data

The next section will introduce the major contribution of this manuscript. The author will propose a new technique based on negative database and the deception of bad users or hackers. The proposed technique is special negative database. This manuscript contribution can be summarized to enhance preserving privacy in big data and avoid time consuming, losing data integrity, complexity and violating any other big data challenges like its size, fault tolerance, timeliness and scalability.

III. Proposed Technique

In this section, the authors introduce a new technique to protect privacy in big data in a new manner that deceives bad users or hackers.

Bad user cannot differentiate between the original data and the data after applying this technique. This technique is called Special Negative Database (SNDB). SNDB is based on deceiving bad users and hackers by replacing only sensitive attribute with its complement. SNDB takes into consideration all attribute types such as binomial, numeric, polynomial as mentioned in Fig. 1. The authors divided the technique into different cases as we will see in the following subsections.

A. Binomial

Binomial attribute means having only two values. Binomial consists of two categories one of them is binary and the other is Boolean. Binary consists of two values 0 or 1 but Boolean is like True or False values or other two values that are vice versa. SNDB deals with binomial attribute by replacing the value with its complement.

B. Numeric

The numeric attribute can be either integer or real numbers. SNDB deals with the numeric value in a different manner. It computes the complement of the digit into 9 individually with a maximum of 4 digits from right taking into consideration the national ID.

For example, if the numeric value is 2896547, the value after complement will be 2893452. While in real numbers, SNDB computes the complement of the digit into 9 individually with a maximum of 4 digits from left.

C. Date and Time

In the case of the date attribute, the date is divided into the year, month, and day. If the value represents a year, SNDB complement each value as a whole to current year when values are less than current year. If some values are equal to current year, the complement of each value will be a whole to Current year+1. If the value represents a month, SNDB will compute the complement of the value into 13. If the value represents a day, SNDB will compute the complement of the digit into 31.

In the case of time attribute. The time is divided into hours, minutes, and seconds. If the value represents hours, SNDB will compute the complement of the value into 24. If the value represents minutes or seconds, SNDB will compute the complement of the value into 9 for the right digit and into 6 for the left digit.

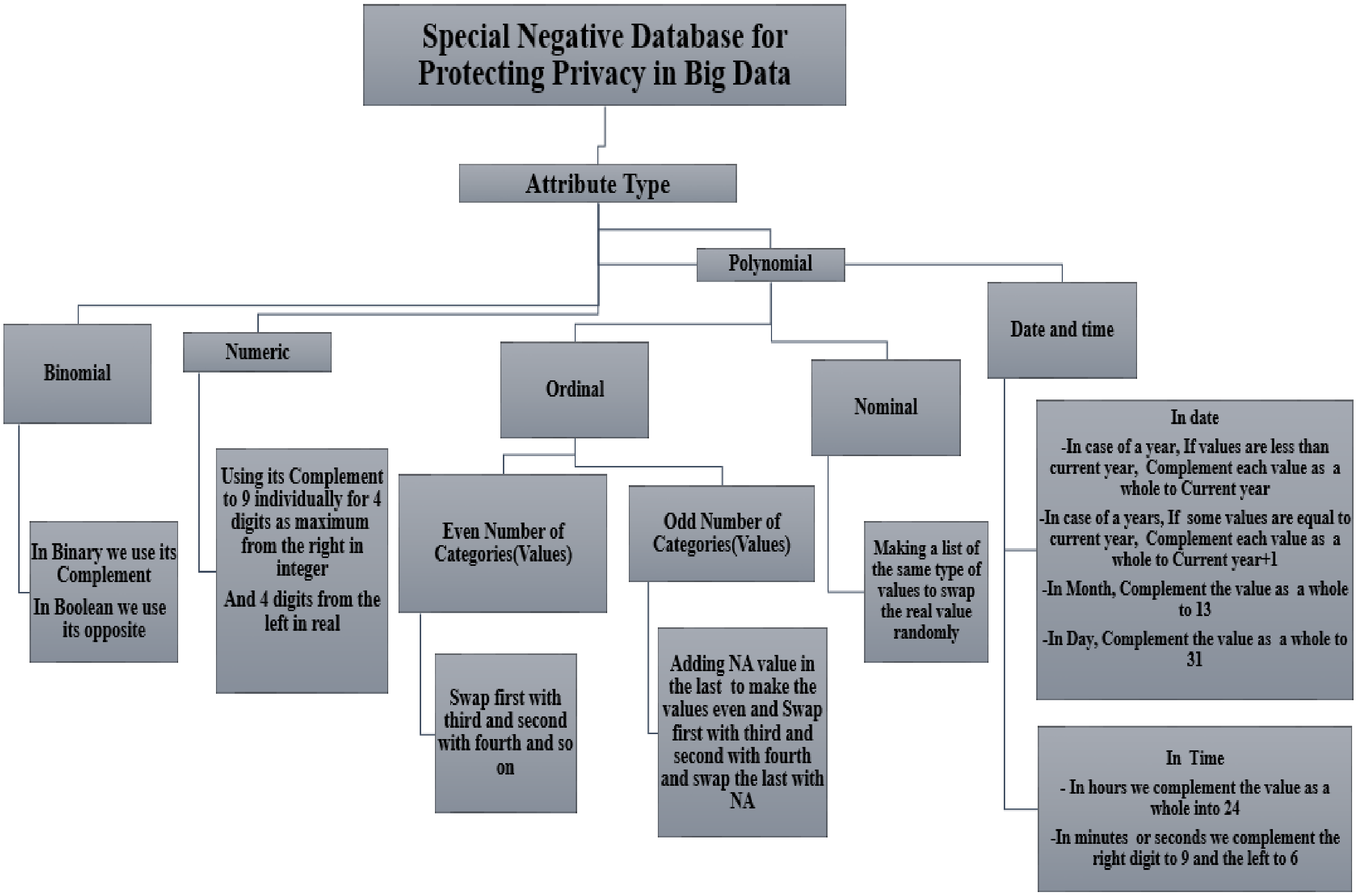

Special Negative Database for Protecting Privacy in Big Data
Attribute Type
Polynomial
Binomial
Numeric
Ordinal
Nominal
Date and time
In Binary we use its Complement
In Boolean we use its opposite
Using its Complement to 9 individually for 4 digits as maximum from the right in integer
And 4 digits from the left in real
Even Number of Categories(Values)
Odd Number of Categories(Values)
Swap first with third and second with fourth and so on
Adding NA value in the last to make the values even and Swap first with third and second with fourth and swap the last with NA
Making a list of the same type of values to swap the real value randomly
In date
-In case of a year, If values are less than current year, Complement each value as a whole to Current year
-In case of a years, If some values are equal to current year, Complement each value as a whole to Current year+1
-In Month, Complement the value as a whole to 13
-In Day, Complement the value as a whole to 31
In Time
- In hours we complement the value as a whole into 24
-In minutes or seconds we complement the right digit to 9 and the left to 6

Fig. 1. Architecture of Special Negative Database (SNDB).

D. Polynomial

Polynomial means that the attribute has more than two text values. Polynomial is divided into two main types. The first is called ordinal and the second is called nominal.

1) Ordinal: Ordinal means that the values can be categorized, classified, ranked, and ordered. Drink size, for example, is an ordinal feature that correlates to the sizes of drinks available at fast-food restaurants. Small, medium, and large are the three possible values for this nominal attribute. These values have a logical order (which correlates to increasing drink size), but the values do not indicate how much larger a medium is than a large. Grade (e.g., A+, A, A-, B+, and so on) and professional rank are two further examples of ordinal characteristics [18].

In ordinal, the number of categories is very important in dealing with the swapping technique. Swapping is used to deceive the bad user that the data is real. In the case of an even number of categories, swapping is very easy to be implemented. SNDB will swap the first value with the third and the second with the fourth and so on. But in case of the odd number of categories, first, we will add the not allowed value (NA) at the last then swapping will be applied as an even manner.

2) Nominal: A nominal attribute's values are names of things or symbols. Nominal implies "related to names." Because each value reflects a state, code or category, nominal characteristics are also known as categorical. There is no discernible order to the values.

Enumerations are another term for values in computer science. Hair color and marital status are two attributes are examples of nominal. Black, blond, red, brown, auburn, grey, and white are all conceivable hair color values in our system. The value of single, married, divorced, or widowed can be assigned to the characteristic of marital status. Both marital status and hair color are also examples of nominal attributes. An occupation is another example of a nominal attribute, having values such as teacher, programmer, farmer dentist, and so on [18].

In the Nominal case, it's important to make a list of values to swap between the real value and one from this list randomly. If the values are names of persons, a list of names will be created. Then one value from this list will be selected and replaced randomly with the original one saving its index in the list. This operation will be repeated again and so on. Using the index, we can get the original data.

IV. DATASETS

In this paper, we apply our model on different datasets according to sensitive attributes taking into consideration all data types.

A. Pollution Dataset

This dataset is about pollution in the United States. The EPA has well-documented pollution in the United States, but downloading all of the data and arranging it in a format that data scientists are interested in is a nuisance. As a result, I gathered data for four key pollutants (nitrogen dioxide, carbon monoxide, ozone, and Sulphur dioxide) for every day between 2000 and 2016 and organized them in a CSV file. There are a total of twenty-eight fields. Each of the four pollutants (NO2, CO, O3, and SO2) has its own set of five columns. 1746661 observations were made. The city, date local, and CO mean are all sensitive parameters.

B. Prouni

This dataset is about Brazil student's scholarship given by Brazilian government on the Prouni program. It contains data from 2005 to 2019 and each line of it corresponds to a student who benefits or has benefited from the Prouni program along with details about them. This dataset consists of 2692540 records.

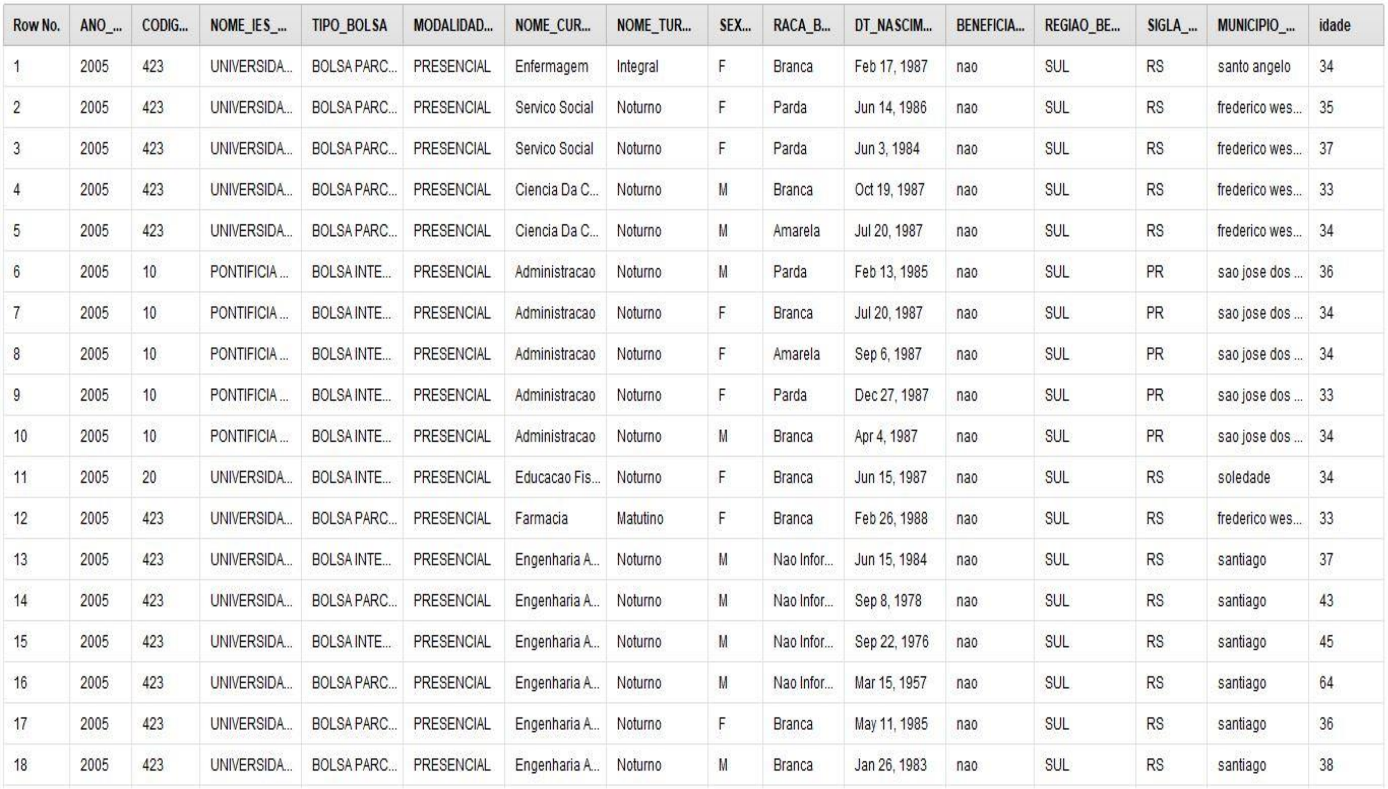
V. RESULTS AND EVALUATIONS

This section will list the results and the evaluation of applying SNDB technique on the datasets for privacy preserving. The author of this paper introduces results for applying SNDB on sensitive attributes in case of binomial, year date and full date and the rest of attribute types will be introduced in the next paper.

A. Binomial Results

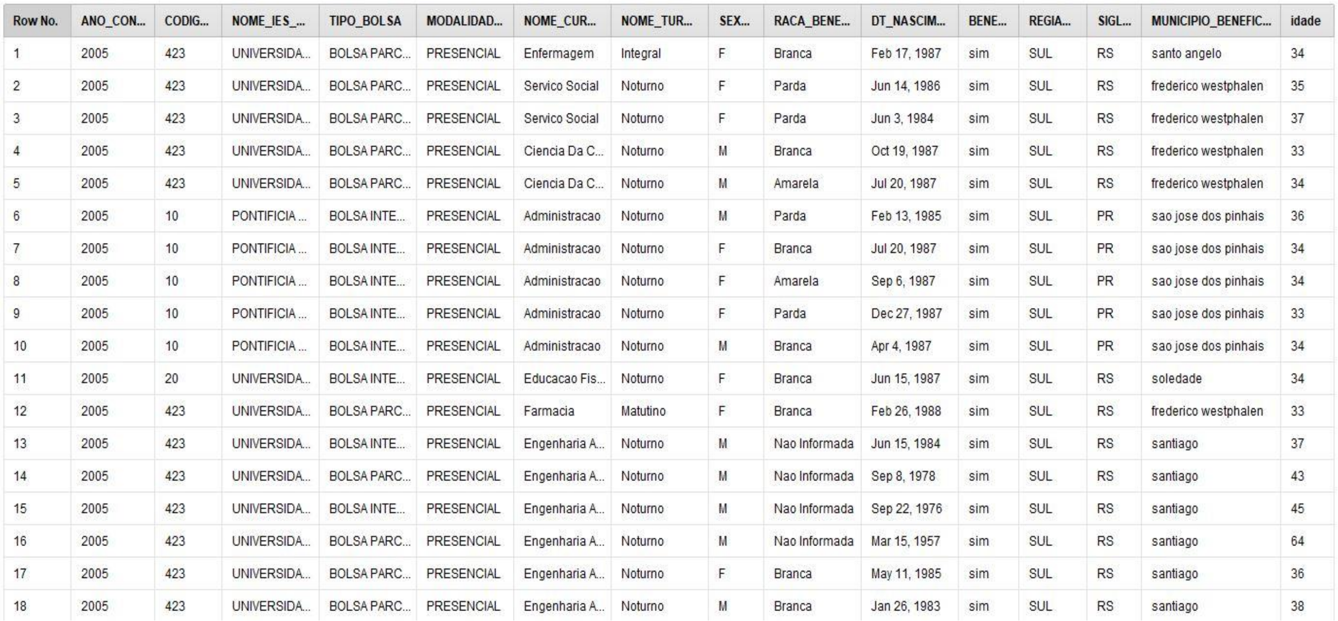
Assuming that the sensitive attribute is BENEFICIARIO_DEFICIENTE_FISICO that means does student have special needs. SNDB swap the value nao that means no with the value sim that means yes and vice versa. Fig. 2 and Fig. 3 illustrate one sample of Prouni dataset before applying SNDB technique and another sample after applying it.

Fig. 4 shows computing some statistical operations on Prouni dataset before applying SNDB technique to get the type of scholarship according to special need. The results show that the number of students who have a special need and have got BOLSA PARCIAL 50% scholarship is 4,947 while the number of who do not have a special need with the same scholarship is 808,557. But students with special needs who have got BOLSA INTEGRAL scholarship is 14,222 while the number of the students who do not have special needs with the same scholarship is 1,862,484. On the other hand, the number of the students with special needs who have got BOLSA COMPLEMENTAR 25% scholarship is 4. While the number of who do not have special needs with the same scholarship is 2,326.



Row No.	ANO_...	CODIG...	NOME_IES_...	TIPO_BOLSA	MODALIDAD...	NOME_CUR...	NOME_TUR...	SEX...	RACA_B...	DT_NASCIM...	BENEFICIA...	REGIAO_BE...	SIGLA_...	MUNICIPIO_...	idade
1	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Enfermagem	Integral	F	Branca	Feb 17, 1987	nao	SUL	RS	santo angelo	34
2	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Servico Social	Noturno	F	Parda	Jun 14, 1986	nao	SUL	RS	frederico wes...	35
3	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Servico Social	Noturno	F	Parda	Jun 3, 1984	nao	SUL	RS	frederico wes...	37
4	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Ciencia Da C...	Noturno	M	Branca	Oct 19, 1987	nao	SUL	RS	frederico wes...	33
5	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Ciencia Da C...	Noturno	M	Amarela	Jul 20, 1987	nao	SUL	RS	frederico wes...	34
6	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	M	Parda	Feb 13, 1985	nao	SUL	PR	sao jose dos ...	36
7	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	F	Branca	Jul 20, 1987	nao	SUL	PR	sao jose dos ...	34
8	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	F	Amarela	Sep 6, 1987	nao	SUL	PR	sao jose dos ...	34
9	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	F	Parda	Dec 27, 1987	nao	SUL	PR	sao jose dos ...	33
10	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	M	Branca	Apr 4, 1987	nao	SUL	PR	sao jose dos ...	34
11	2005	20	UNIVERSIDA...	BOLSA INTE...	PRESENCIAL	Educacao Fis...	Noturno	F	Branca	Jun 15, 1987	nao	SUL	RS	soledade	34
12	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Farmacia	Matutino	F	Branca	Feb 26, 1988	nao	SUL	RS	frederico wes...	33
13	2005	423	UNIVERSIDA...	BOLSA INTE...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Infor...	Jun 15, 1984	nao	SUL	RS	santiago	37
14	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Infor...	Sep 8, 1978	nao	SUL	RS	santiago	43
15	2005	423	UNIVERSIDA...	BOLSA INTE...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Infor...	Sep 22, 1976	nao	SUL	RS	santiago	45
16	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Infor...	Mar 15, 1957	nao	SUL	RS	santiago	64
17	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	F	Branca	May 11, 1985	nao	SUL	RS	santiago	36
18	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	M	Branca	Jan 26, 1983	nao	SUL	RS	santiago	38

Fig. 2. Sample of Prouni Dataset before Applying SNDB.



Row No.	ANO_CON...	CODIG...	NOME_IES_...	TIPO_BOLSA	MODALIDAD...	NOME_CUR...	NOME_TUR...	SEX...	RACA_BENE...	DT_NASCIM...	BENE...	REGIA...	SIGL...	MUNICIPIO_BENEFIC...	idade
1	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Enfermagem	Integral	F	Branca	Feb 17, 1987	sim	SUL	RS	santo angelo	34
2	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Servico Social	Noturno	F	Parda	Jun 14, 1986	sim	SUL	RS	frederico westphalen	35
3	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Servico Social	Noturno	F	Parda	Jun 3, 1984	sim	SUL	RS	frederico westphalen	37
4	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Ciencia Da C...	Noturno	M	Branca	Oct 19, 1987	sim	SUL	RS	frederico westphalen	33
5	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Ciencia Da C...	Noturno	M	Amarela	Jul 20, 1987	sim	SUL	RS	frederico westphalen	34
6	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	M	Parda	Feb 13, 1985	sim	SUL	PR	sao jose dos pinhais	36
7	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	F	Branca	Jul 20, 1987	sim	SUL	PR	sao jose dos pinhais	34
8	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	F	Amarela	Sep 6, 1987	sim	SUL	PR	sao jose dos pinhais	34
9	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	F	Parda	Dec 27, 1987	sim	SUL	PR	sao jose dos pinhais	33
10	2005	10	PONTIFICIA ...	BOLSA INTE...	PRESENCIAL	Administracao	Noturno	M	Branca	Apr 4, 1987	sim	SUL	PR	sao jose dos pinhais	34
11	2005	20	UNIVERSIDA...	BOLSA INTE...	PRESENCIAL	Educacao Fis...	Noturno	F	Branca	Jun 15, 1987	sim	SUL	RS	soledade	34
12	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Farmacia	Matutino	F	Branca	Feb 26, 1988	sim	SUL	RS	frederico westphalen	33
13	2005	423	UNIVERSIDA...	BOLSA INTE...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Informada	Jun 15, 1984	sim	SUL	RS	santiago	37
14	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Informada	Sep 8, 1978	sim	SUL	RS	santiago	43
15	2005	423	UNIVERSIDA...	BOLSA INTE...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Informada	Sep 22, 1976	sim	SUL	RS	santiago	45
16	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	M	Nao Informada	Mar 15, 1957	sim	SUL	RS	santiago	64
17	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	F	Branca	May 11, 1985	sim	SUL	RS	santiago	36
18	2005	423	UNIVERSIDA...	BOLSA PARC...	PRESENCIAL	Engenharia A...	Noturno	M	Branca	Jan 26, 1983	sim	SUL	RS	santiago	38

Fig. 3. Sample of Prouni Dataset after Applying SNDB.

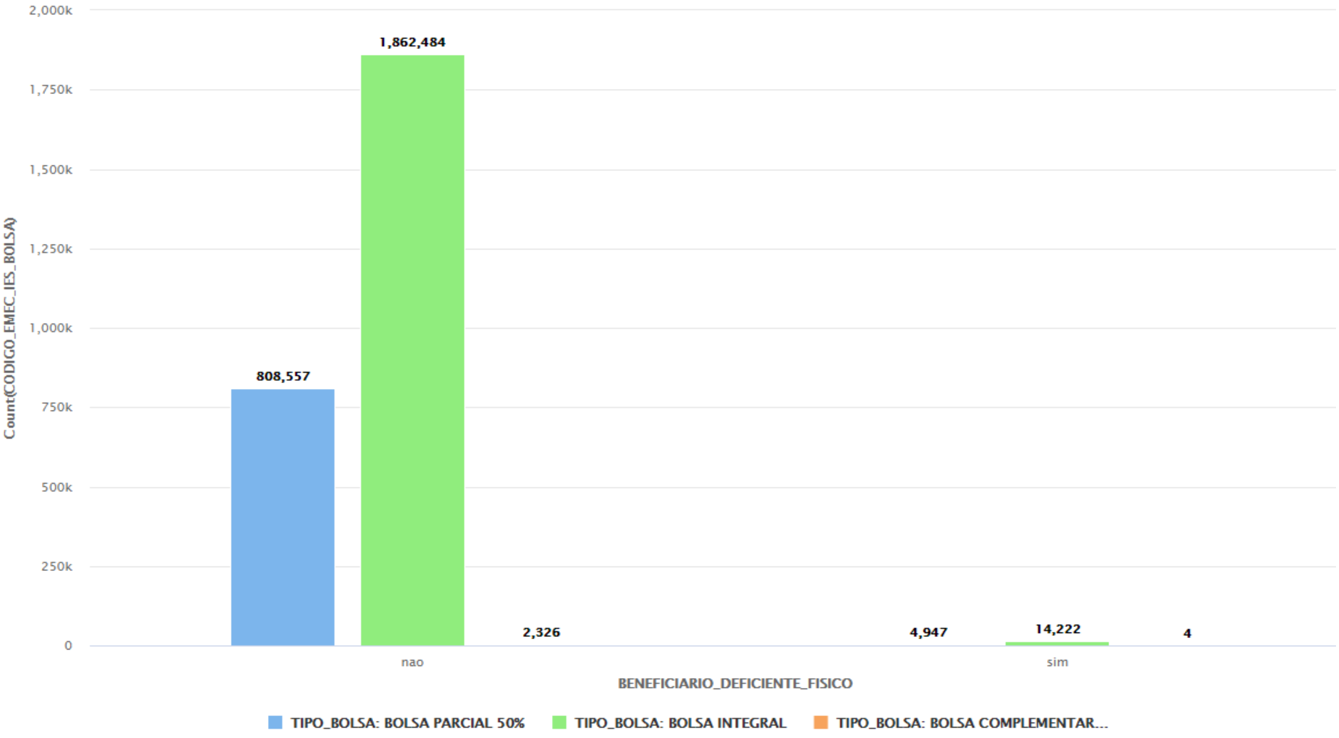


Fig. 4. Type of Scholarship with respect to Special need before Applying SNDB.

Fig. 5 shows computing some statistical operations on Prouni dataset after applying SNDB technique to get the type of scholarship according to special needs. The results show that the number of students who have a special need and have got BOLSA PARCIAL 50% scholarship is 808,557 while the number of who do not have special needs with the same scholarship is 4,947. But students with special needs who have got BOLSA INTEGRAL scholarship is 1,862,484 while the number of who do not have special needs with the same scholarship is 14,222. On the other hand, the number of the students with special needs who have got BOLSA COMPLEMENTAR 25% scholarship is 2,326. While the number of students who do not have special needs with the same scholarship is 4. Fig. 4 and Fig. 5 show that the results have a big difference before applying SNDB and after applying SNDB on the same dataset which mean the success of the algorithm when applying on binomial sensitive attribute.

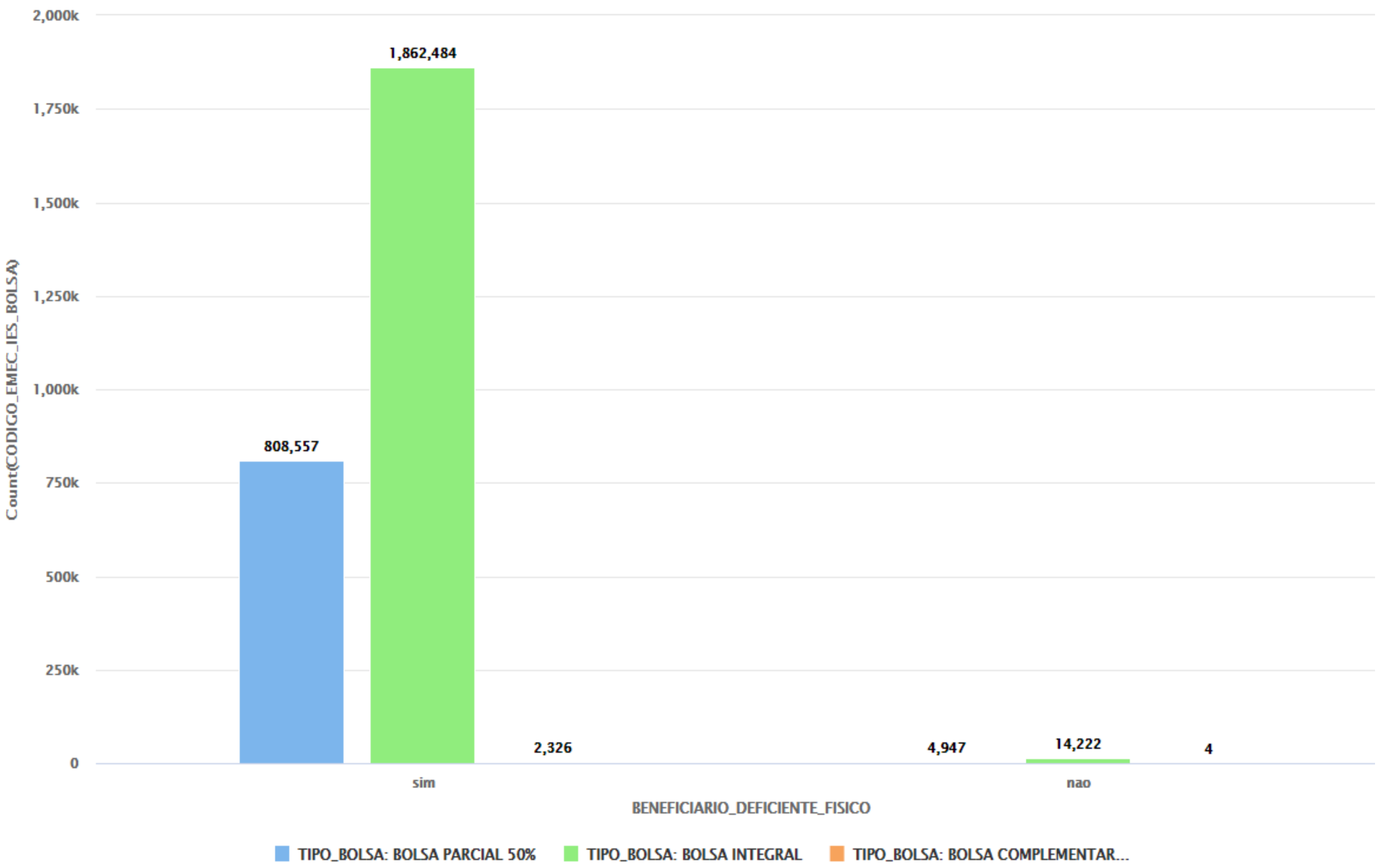


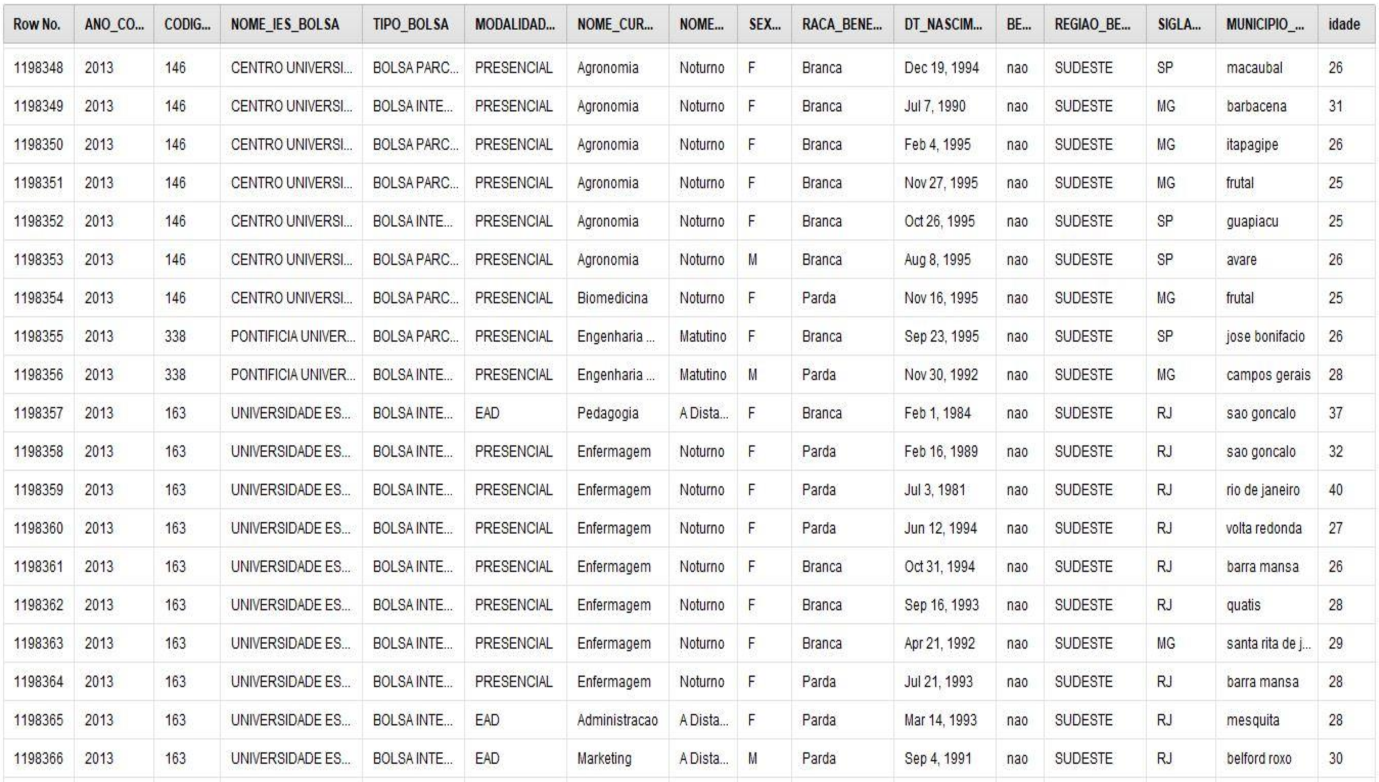
Fig. 5. Type of Scholarship with respect to Special Need after Applying SNDB.

B. Date Results

1) Year date: SNDB technique deals with sensitive attribute with the type date in a special manner. Assuming the sensitive attribute is ANO_CONCESSAO_BOLSA that means the year of the scholarship. It consists of a year only. SNDB applies the complement of a year to the current year (2021). Fig. 6 and Fig. 7 illustrate one sample of Prouni dataset before applying SNDB technique and another sample after applying SNDB on ANO_CONCESSAO_BOLSA.

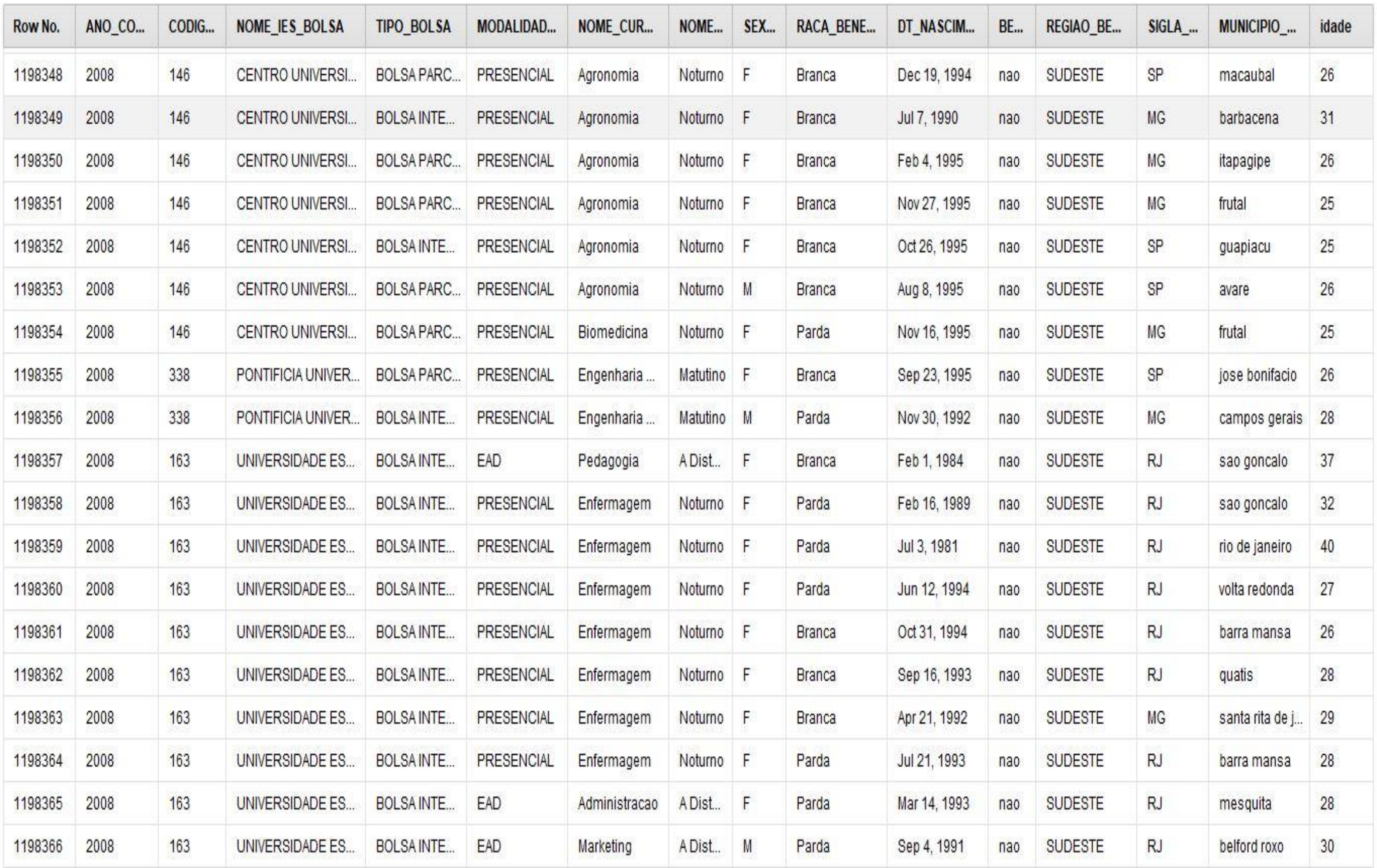
Fig. 8 and Fig. 9 illustrate the difference between gender of the student who has got the scholarship before applying SNDB and after applying it. If we check out Fig. 8 and Fig. 9, we will say that the years of scholarship in the original data are from 2005 to 2019 while years from 2002 to 2016 are the years of scholarship after applying SNDB technique. When taking 2005 as an example, we will see the number of male students is 36,097 and the number of female students is 39,532 in the original data. While in the data after applying SNDB, the number of male students is 108,057 and the number of female students is 131,205 in original data. Fig. 8 and Fig. 9 show that the results have a big difference before applying SNDB and after applying SNDB on ANO_CONCESSAO_BOLSA in the same dataset which means the success of the algorithm when applying on date sensitive attribute of year value only. In the next section the paper will show the result of applying SNDB on different date value.

Full date: Assuming the sensitive attribute is date local. SNDB deals with date local in a different manner, SNDB changes the month only because changing the month is sufficient to change the original data. SNDB swaps January with December and vice versa, February with November and vice versa, March with October and vice versa, April with September and vice versa, May with August and vice versa and Jun with July and vice versa. Fig. 10 and Fig. 11 illustrate one sample of pollution dataset before applying SNDB technique and another sample after applying SNDB on date local attribute. Fig. 12 and Fig. 13 below illustrate the difference between the maximum value of (Sulphur Dioxide and Nitrogen Dioxide) mean before and after applying SNDB on date local. This paper takes the first five days of December,2000 as an example to show the difference between the original dataset and the dataset after applying SNDB technique. When checking out Fig. 12 and 13, we will see the big difference between the values of original and SNDB dataset.



Row No.	ANO_CO...	CODIG...	NOME_IES_BOLSA	TIPO_BOLSA	MODALIDAD...	NOME_CUR...	NOME...	SEX...	RACA_BENE...	DT_NASCIM...	BE...	REGIAO_BE...	SIGLA...	MUNICIPIO_...	idade
1198348	2013	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	F	Branca	Dec 19, 1994	nao	SUDESTE	SP	macaubal	26
1198349	2013	146	CENTRO UNIVERSI...	BOLSA INTE...	PRESENCIAL	Agronomia	Noturno	F	Branca	Jul 7, 1990	nao	SUDESTE	MG	barbacena	31
1198350	2013	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	F	Branca	Feb 4, 1995	nao	SUDESTE	MG	itapagipe	26
1198351	2013	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	F	Branca	Nov 27, 1995	nao	SUDESTE	MG	frutal	25
1198352	2013	146	CENTRO UNIVERSI...	BOLSA INTE...	PRESENCIAL	Agronomia	Noturno	F	Branca	Oct 26, 1995	nao	SUDESTE	SP	guapiacu	25
1198353	2013	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	M	Branca	Aug 8, 1995	nao	SUDESTE	SP	avare	26
1198354	2013	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Biomedicina	Noturno	F	Parda	Nov 16, 1995	nao	SUDESTE	MG	frutal	25
1198355	2013	338	PONTIFICIA UNIVER...	BOLSA PARC...	PRESENCIAL	Engenharia ...	Matutino	F	Branca	Sep 23, 1995	nao	SUDESTE	SP	jose bonifacio	26
1198356	2013	338	PONTIFICIA UNIVER...	BOLSA INTE...	PRESENCIAL	Engenharia ...	Matutino	M	Parda	Nov 30, 1992	nao	SUDESTE	MG	campos gerais	28
1198357	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	EAD	Pedagogia	A Dista...	F	Branca	Feb 1, 1984	nao	SUDESTE	RJ	sao goncalo	37
1198358	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Feb 16, 1989	nao	SUDESTE	RJ	sao goncalo	32
1198359	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Jul 3, 1981	nao	SUDESTE	RJ	rio de janeiro	40
1198360	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Jun 12, 1994	nao	SUDESTE	RJ	volta redonda	27
1198361	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Branca	Oct 31, 1994	nao	SUDESTE	RJ	barra mansa	26
1198362	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Branca	Sep 16, 1993	nao	SUDESTE	RJ	quatis	28
1198363	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Branca	Apr 21, 1992	nao	SUDESTE	MG	santa rita de j...	29
1198364	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Jul 21, 1993	nao	SUDESTE	RJ	barra mansa	28
1198365	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	EAD	Administracao	A Dista...	F	Parda	Mar 14, 1993	nao	SUDESTE	RJ	mesquita	28
1198366	2013	163	UNIVERSIDADE ES...	BOLSA INTE...	EAD	Marketing	A Dista...	M	Parda	Sep 4, 1991	nao	SUDESTE	RJ	belford roxo	30

Fig. 6. Sample of Prouni Dataset before Applying SNDB on ano_concessao_bolsa.



Row No.	ANO_CO...	CODIG...	NOME_IES_BOLSA	TIPO_BOLSA	MODALIDAD...	NOME_CUR...	NOME...	SEX...	RACA_BENE...	DT_NASCIM...	BE...	REGIAO_BE...	SIGLA_...	MUNICIPIO_...	idade
1198348	2008	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	F	Branca	Dec 19, 1994	nao	SUDESTE	SP	macaubal	26
1198349	2008	146	CENTRO UNIVERSI...	BOLSA INTE...	PRESENCIAL	Agronomia	Noturno	F	Branca	Jul 7, 1990	nao	SUDESTE	MG	barbacena	31
1198350	2008	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	F	Branca	Feb 4, 1995	nao	SUDESTE	MG	itapagipe	26
1198351	2008	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	F	Branca	Nov 27, 1995	nao	SUDESTE	MG	frutal	25
1198352	2008	146	CENTRO UNIVERSI...	BOLSA INTE...	PRESENCIAL	Agronomia	Noturno	F	Branca	Oct 26, 1995	nao	SUDESTE	SP	guapiacu	25
1198353	2008	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Agronomia	Noturno	M	Branca	Aug 8, 1995	nao	SUDESTE	SP	avare	26
1198354	2008	146	CENTRO UNIVERSI...	BOLSA PARC...	PRESENCIAL	Biomedicina	Noturno	F	Parda	Nov 16, 1995	nao	SUDESTE	MG	frutal	25
1198355	2008	338	PONTIFICIA UNIVER...	BOLSA PARC...	PRESENCIAL	Engenharia ...	Matutino	F	Branca	Sep 23, 1995	nao	SUDESTE	SP	jose bonifacio	26
1198356	2008	338	PONTIFICIA UNIVER...	BOLSA INTE...	PRESENCIAL	Engenharia ...	Matutino	M	Parda	Nov 30, 1992	nao	SUDESTE	MG	campos gerais	28
1198357	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	EAD	Pedagogia	A Dist...	F	Branca	Feb 1, 1984	nao	SUDESTE	RJ	sao goncalo	37
1198358	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Feb 16, 1989	nao	SUDESTE	RJ	sao goncalo	32
1198359	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Jul 3, 1981	nao	SUDESTE	RJ	rio de janeiro	40
1198360	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Jun 12, 1994	nao	SUDESTE	RJ	volta redonda	27
1198361	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Branca	Oct 31, 1994	nao	SUDESTE	RJ	barra mansa	26
1198362	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Branca	Sep 16, 1993	nao	SUDESTE	RJ	quatis	28
1198363	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Branca	Apr 21, 1992	nao	SUDESTE	MG	santa rita de j...	29
1198364	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	PRESENCIAL	Enfermagem	Noturno	F	Parda	Jul 21, 1993	nao	SUDESTE	RJ	barra mansa	28
1198365	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	EAD	Administracao	A Dist...	F	Parda	Mar 14, 1993	nao	SUDESTE	RJ	mesquita	28
1198366	2008	163	UNIVERSIDADE ES...	BOLSA INTE...	EAD	Marketing	A Dist...	M	Parda	Sep 4, 1991	nao	SUDESTE	RJ	belford roxo	30

Fig. 7. Sample of Prouni Dataset after Applying SNDB on ano_concessao_bolsa.

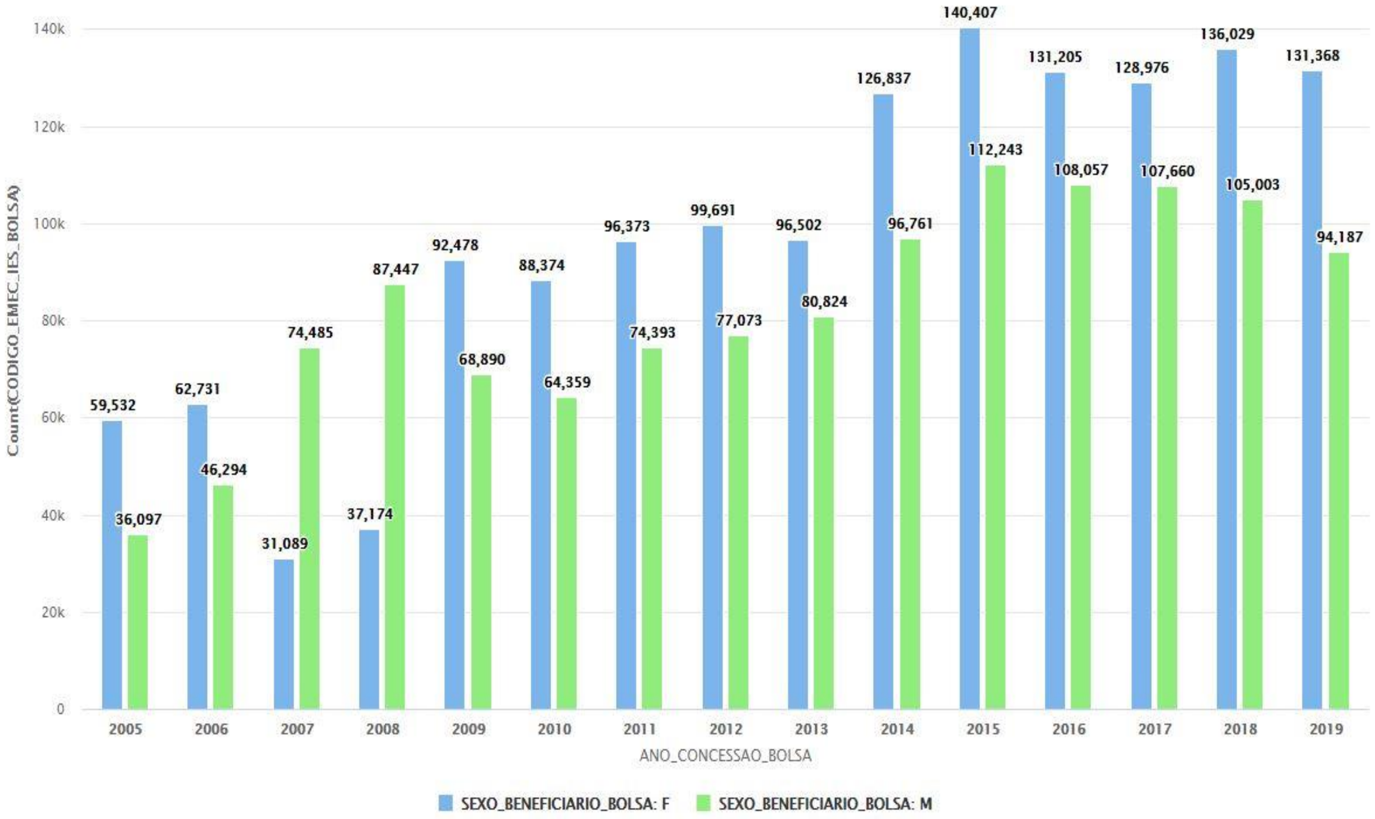


Fig. 8. Gender of Students in the Year of Scholarship before Applying SNDB.

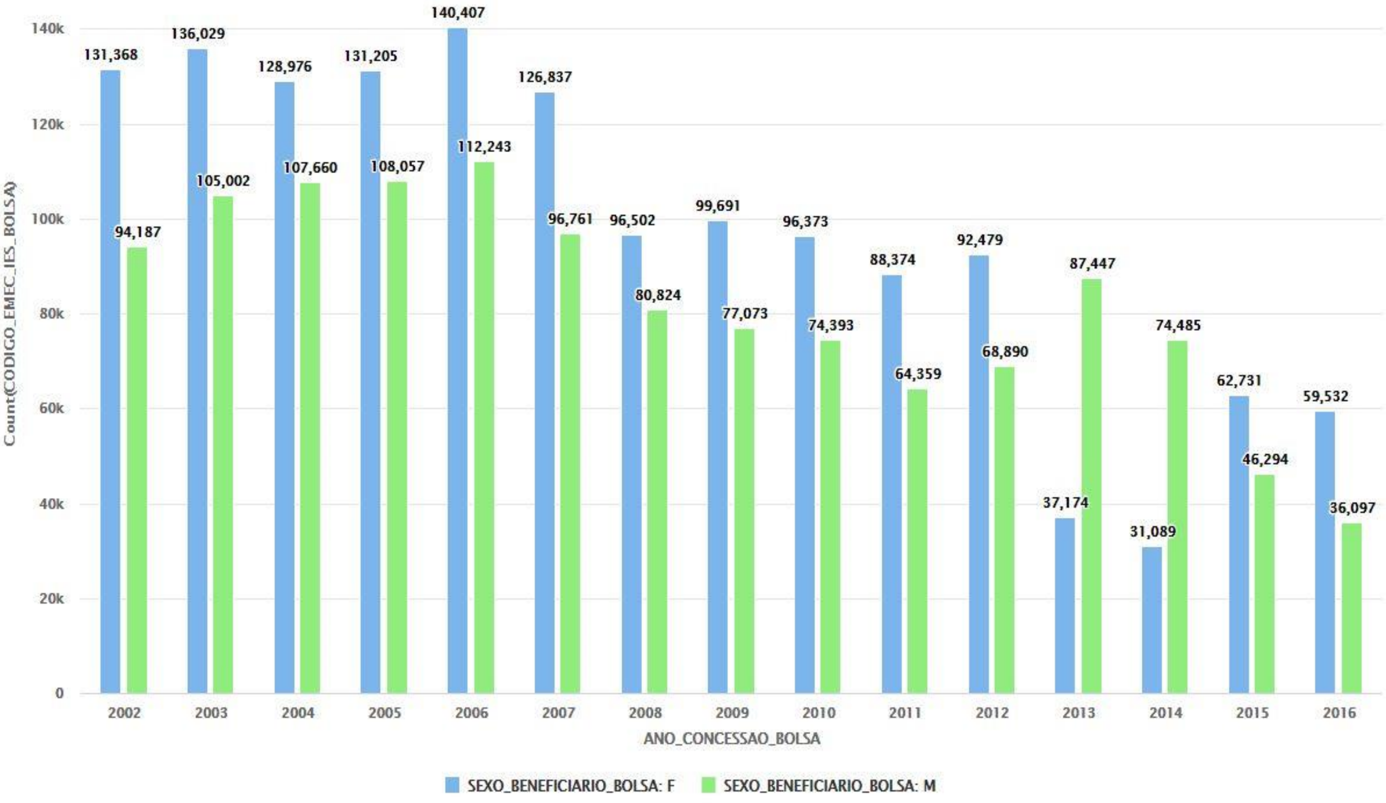
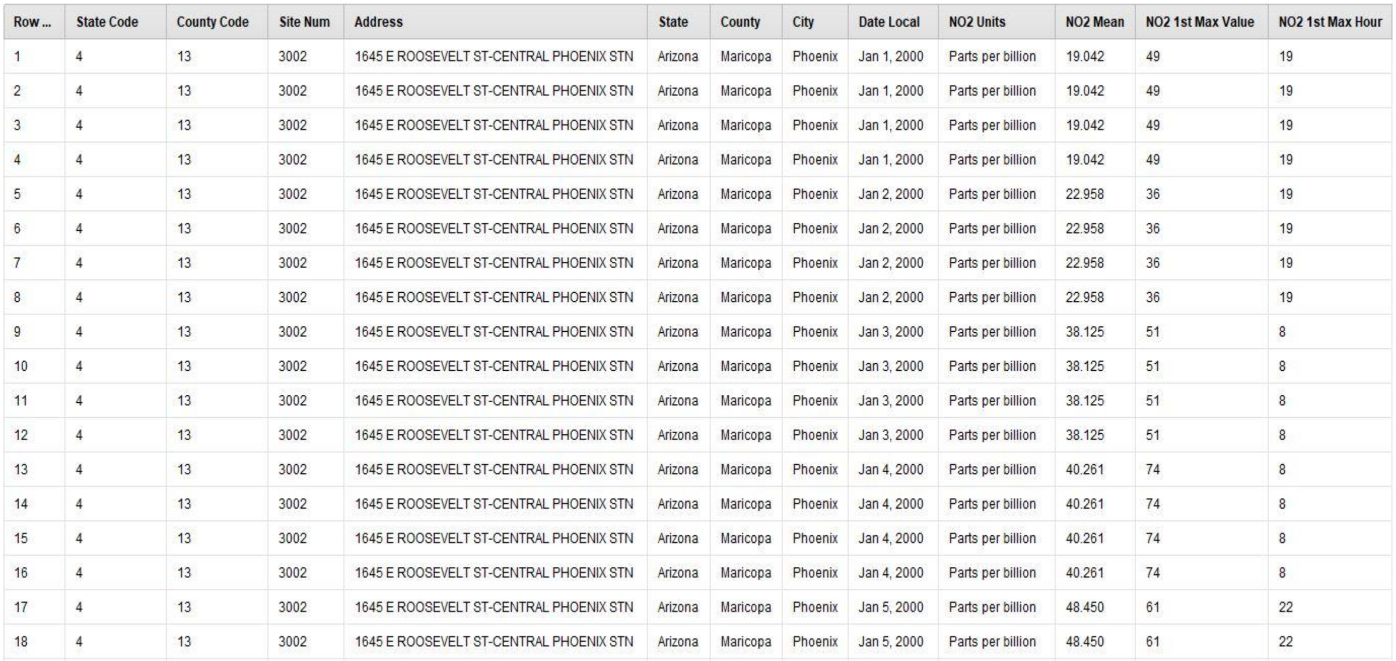
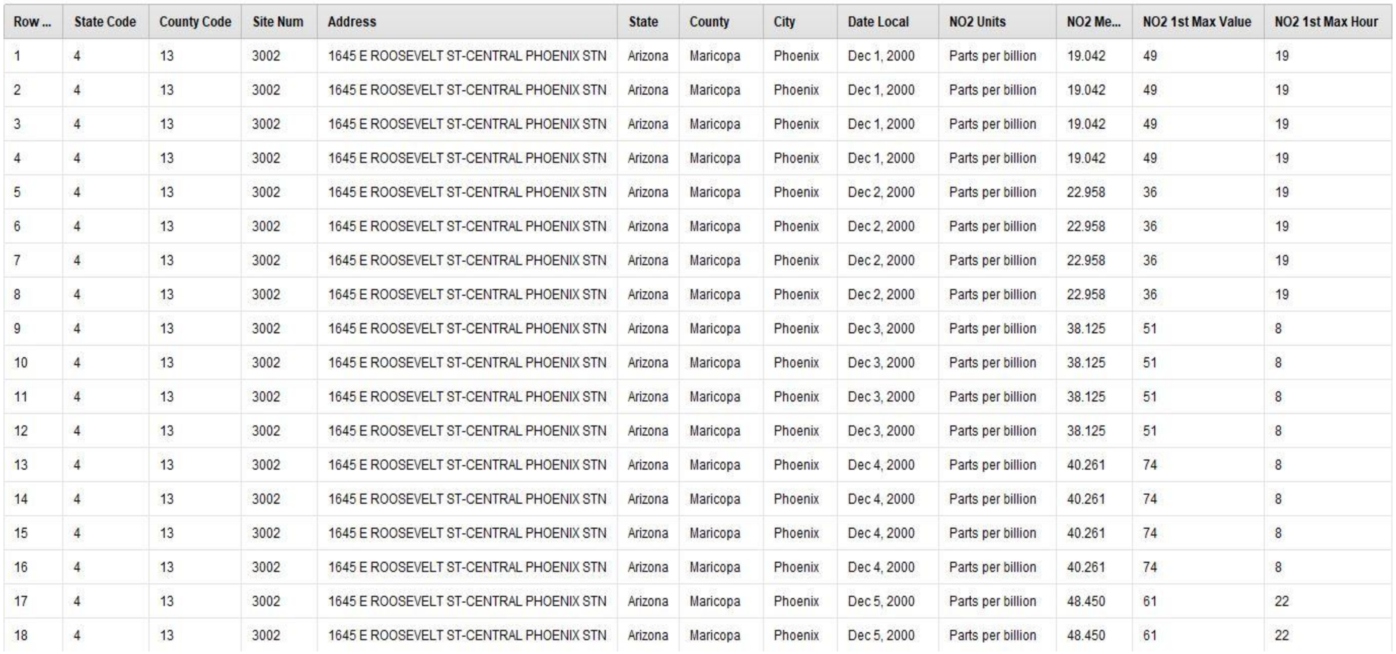


Fig. 9. Gender of Students in the Year of Scholarship after Applying SNDB.



Row ...	State Code	County Code	Site Num	Address	State	County	City	Date Local	NO2 Units	NO2 Mean	NO2 1st Max Value	NO2 1st Max Hour
1	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 1, 2000	Parts per billion	19.042	49	19
2	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 1, 2000	Parts per billion	19.042	49	19
3	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 1, 2000	Parts per billion	19.042	49	19
4	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 1, 2000	Parts per billion	19.042	49	19
5	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 2, 2000	Parts per billion	22.958	36	19
6	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 2, 2000	Parts per billion	22.958	36	19
7	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 2, 2000	Parts per billion	22.958	36	19
8	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 2, 2000	Parts per billion	22.958	36	19
9	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 3, 2000	Parts per billion	38.125	51	8
10	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 3, 2000	Parts per billion	38.125	51	8
11	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 3, 2000	Parts per billion	38.125	51	8
12	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 3, 2000	Parts per billion	38.125	51	8
13	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 4, 2000	Parts per billion	40.261	74	8
14	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 4, 2000	Parts per billion	40.261	74	8
15	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 4, 2000	Parts per billion	40.261	74	8
16	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 4, 2000	Parts per billion	40.261	74	8
17	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 5, 2000	Parts per billion	48.450	61	22
18	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Jan 5, 2000	Parts per billion	48.450	61	22

Fig. 10. Sample of Pollution Dataset before Applying SNDB on Date Local.



Row ...	State Code	County Code	Site Num	Address	State	County	City	Date Local	NO2 Units	NO2 Me...	NO2 1st Max Value	NO2 1st Max Hour
1	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 1, 2000	Parts per billion	19.042	49	19
2	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 1, 2000	Parts per billion	19.042	49	19
3	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 1, 2000	Parts per billion	19.042	49	19
4	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 1, 2000	Parts per billion	19.042	49	19
5	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 2, 2000	Parts per billion	22.958	36	19
6	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 2, 2000	Parts per billion	22.958	36	19
7	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 2, 2000	Parts per billion	22.958	36	19
8	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 2, 2000	Parts per billion	22.958	36	19
9	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 3, 2000	Parts per billion	38.125	51	8
10	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 3, 2000	Parts per billion	38.125	51	8
11	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 3, 2000	Parts per billion	38.125	51	8
12	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 3, 2000	Parts per billion	38.125	51	8
13	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 4, 2000	Parts per billion	40.261	74	8
14	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 4, 2000	Parts per billion	40.261	74	8
15	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 4, 2000	Parts per billion	40.261	74	8
16	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 4, 2000	Parts per billion	40.261	74	8
17	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 5, 2000	Parts per billion	48.450	61	22
18	4	13	3002	1645 E ROOSEVELT ST-CENTRAL PHOENIX STN	Arizona	Maricopa	Phoenix	Dec 5, 2000	Parts per billion	48.450	61	22

Fig. 11. Sample of Pollution Dataset after Applying SNDB on Date Local.

As seen from results figures above, SNDB is valid for big data since the size of data does not change before and after applying SNDB. Processing time is fast when comparing to traditional negative database. The deception of SNDB technique is big and this makes privacy level is stronger. Bad users or hackers cannot differentiate between the original data and the data after applying SNDB technique. This makes the decryption very hard for bad users while it is very easy for data owner to decrypt the SNDB data. Table I below shows the comparison between SNDB and traditional negative database. There are comparative results on different datasets because SNDB technique is applied according to sensitive attributes covering all data types of each dataset.

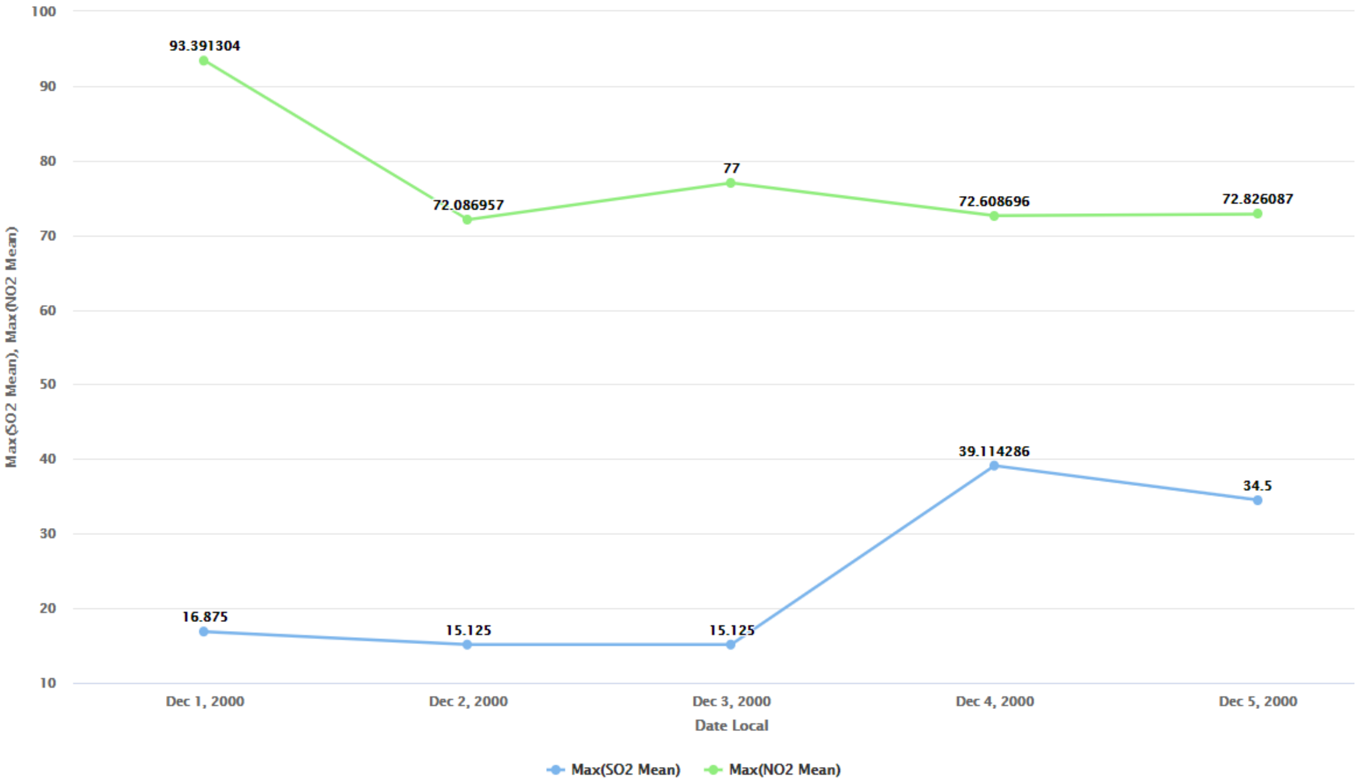


Fig. 12. Maximum Value of (Sulphur Dioxide and Nitrogen Dioxide) mean before Applying SNDB on Date Local.

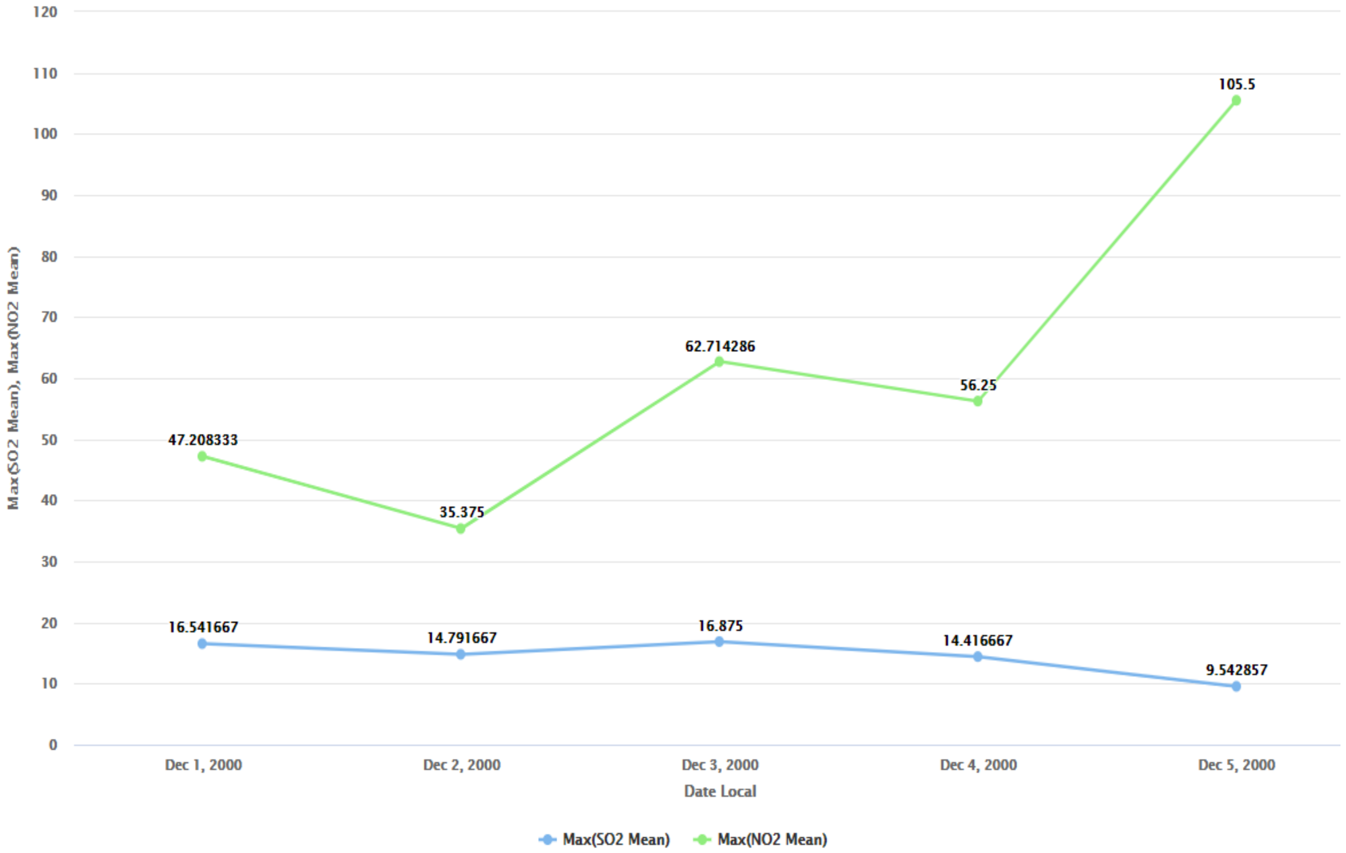


Fig. 13. Maximum Value of (Sulphur Dioxide and Nitrogen Dioxide) mean after Applying SNDB on Date Local.

TABLE I. COMPARISON BETWEEN TRADITIONAL NEGATIVE DATABASE AND SNDB

Technique	**Records in the original pollution dataset**	**Records in the pollution dataset after applying the algorithm**	**Records in the original Prouni dataset**	**Records in the Prouni dataset after applying the algorithm**	**Processing Time**	**Deception level**	**Privacy level**	**Decryption for** users and hackers	**Decrypt-ion for** data owner	**Validity for Big data**
Traditional Negative Database	1,746,661	3,493,322 In case of full date attribute	2,692,540	5,385,080 In case of binomial and year date attributes	Slow 32 seconds in case of Pollution dataset and 47 seconds in case of Prouni dataset	There is a little deception for bad users and hackers	Strong	Hard	Easy	Not valid because the size of data increases and doubles
SNDB	1,746,661	1,746,661 In case of full date attribute	2,692,540	2,692,540 In case of binomial and year date attributes	Fast 9 seconds in case of Pollution dataset and 12 seconds in case of Prouni dataset	There is a big deception for bad users and hackers	More stronger	Harder	Easier	Valid because the size of data does not change

VI. CONCLUSION

This paper lists the most important big data challenges and focuses on privacy challenge; it summaries privacy violation situations. The author also provides a list of the most efficient and popular techniques used to protect data privacy with their advantages and drawbacks. The proposed technique in this paper is SNDB based on negative database in different manner. SNDB is based on deceiving bad users and hackers by replacing only sensitive attribute with its complement. SNDB takes into consideration all attribute types such as binomial, numeric, polynomial. SNDB technique is applied on different datasets according to the type of the sensitive attributes of each dataset. In this technique, bad user cannot differentiate between the original data and the data after applying this technique which enhances the level of privacy.

As seen from results, SNDB can avoid drawbacks of previous techniques since it has the advantage of high privacy protection in big data. SNDB has no time consuming since it deals with sensitive attribute only. It also keeps track of data integrity and data size since there is no decreasing or increasing for any record of data and this advantage makes SNDB very suitable for big data. It also has low complexity since it only replaces sensitive attribute value with its complement. After applying SNDB, we can easily get the original data by applying the complement another time according to the rules of the data owner.

VII. FUTURE WORK

Finally, the author provides the results of applying SNDB on big dataset with binomial, year date and full date sensitive attribute. In the future work, the author will introduce the results of applying SNDB on numeric, ordinal and nominal sensitive attributes. Also, the author tends to take into consideration transposition techniques instead of replacing values with each other’s.

REFERENCES

[1] P. V. Desai, “A survey on big data applications and challenges,” In Proc. of the Second International Conf. on Inventive Communication and Computational Technologies (ICICCT), IEEE, pp. 737-740, 2018.

[2] M. M. Shendi, H. M. Elkadi, and M. H. Khafagy, “A study on the big data log analysis: goals, challenges, issues, and tools,” International Journal of Artificial Intelligence and Soft Computing, vol. 7, no. 2, pp. 5-12, 2019.

[3] Sk. M. Gouse and G. K. Mohan, “Improving the Performance of Various Privacy Preserving Databases using Hybrid Geometric Data Perturbation Classification Model,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 10, pp. 249-253, 2020.

[4] O. Almutairi and K. Almarhabi, “Investigation of Smart Home Security and Privacy: Consumer Perception in Saudi Arabia,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 4, pp. 614-622, 2021.

[5] K. Vani and B. Srinivas, “Enhanced slicing for privacy-preserving data publishing,” The International Journal of Engineering and Science (IJES), vol. 2, no. 10, pp. 1-4, 2013.

[6] A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua, and S. Guo “Protection of big data privacy,” IEEE access, vol. 4, pp. 821-1834, 2016.

[7] J. A. Shamsi and M. A. Khojaye, “Understanding privacy violations in big data systems,” IT Professional, vol. 20, no. 3, pp. 73-81, 2018.

[8] L. Rajesh and P. Satyanarayana, “Detecting Flooding Attacks in Communication Protocol of Industrial Control Systems,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 1, pp. 396-401, 2020.

[9] A. A. Abi Sen, F. A. Eassa, and K. Jambi, “Preserving privacy of smart cities based on the fog computing,” In Proc. International Conf. on Smart Cities, Infrastructure, Technologies and Applications, Springer, Cham, pp. 185-191, 2017.

[10] P. Jain, M. Gyanchandani and N. Khare, “Enhanced secured map-reduce layer for big data privacy and security,” Journal of Big Data, vol. 6, no. 1, pp. 1-17, 2019.

[11] M. Yamin, Y. Alsaawy, A. B. Alkhodre, and A. A. A. Sen, “An innovative method for preserving privacy in internet of things,” Journal of Sensors, vol. 19, no. 9, pp. 3355, 2019. [Online]. Available: https://doi.org/10.3390/s19153355.

[12] A. Patel, N. Sharma, and M. Eirinaki, “Negative Database for Data Security,” In Proc. International Conf. on Computing, Engineering and Information, IEEE, pp. 67-70, 2009.

[13] C. Egbunike and S. Rajendran, “The Implementation of Negative Database as a Security Technique on a Generic Database System,” In Proc. International Conf. on circuits Power and Computing Technologies (ICCPCT), IEEE, pp. 1-8, 2017.

[14] A. A. A. Sen, F. A. Eassa, K. Jambi, N. M. Bahbouh, S. S. Albouq, and A. Alshanqiti, “Enhanced- blind approach for privacy protection of iot,” 2020 IEEE 7th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, pp. 240-243, 2020.

[15] S. AR and B. G. Banik, “A Comprehensive Study of Blockchain Services: Future of Cryptography,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 10, pp. 279-288, 2020.

[16] H. V. Abhijith and H. S. Rameshbabu, “Secure data transmission framework for internet of things based on oil spill detection application,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 5, pp. 189-195, 2021.

[17] S. Trichni, F. Omary, and M. Bougrine, “New smart encryption approach based on multidimensional analysis tools,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 5, pp. 666-675, 2021.

[18] H. Jiawei, M. Kamber, and P. Jian, Data Mining Concepts and Techniques, (3rd ed), Morgan Kauffman, 2011.