International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT1228529

A Comprehensive Analysis Of Normalization Approaches For Privacy Protection In Data Mining

Surendra Kumar Reddy Koduru
Business Intelligence & Reporting Lead, NC, USA
Email : surendrakoduru.bi@gmail.com

Article Info
Publication Issue :
Volume 8, Issue 5
September-October-2022
Page Number : 144-157
Article History
Accepted: 10 Sep 2022
Published: 27 Sep 2022

ABSTRACT

Data Mining is a fundamental method of extracting large volumes of data sets by unfamiliar patterns. These extracted data can be shared between enterprises to improve their corporate benefits. For the data mining process, sharing such confidential data is very important. It is very important to safeguard such information against unwanted exposure that leads to privacy leakage. In recent days, privacy in various data mining applications has become very important. In order to overcome such problem privacy, data mining techniques are preserved. It provides accurate data mining results without scarifying the original data values and ensures both accuracy and privacy. Data protection is achieved through the analysis of the data using normalization techniques in this proposed work. The approach proposed in comparison the effects of min-max, decimal scaling, and Z-Score normalization techniques. Experimental effects confirmed that the Min-max Standardization technology archived the most precision with minimum data loss.

Keywords : Data Privacy, Data Accuracy, Privacy Preservation, Z-Score Normalization, Normalization, Privacy

I. INTRODUCTION

In every data mining application, which could consist of private, sensitive information, now a day, large volumes of information are collected and analysed regularly [1][2]. Such data may contain health records, criminal records, credit card information, purchasing practises, customer financial information[3]. All this information plays a key role in improving business, governmental organisational benefits, and decision-making by bringing social benefits, such as medical research, improving national safety and reducing crime. But analysing such information preserves the privacy of each and every person with the new challenge[4]. In data mining, privacy is important as sensitive data originates from heterogeneous sources [5,15,16].

In order to preserve such sensitive information about individuals and information leakage as well as the accuracy of the original data protection techniques for data mining are established [6]. Data disturbance is a popular method of data modification in order to preserve the sensitive information contained in the dataset and achieve good data mining results. Sensitive values in a dataset are modified in this data disruption technique before data mining application begins .This paper uses data transformation methods based on normalisation to modify sensitive values in a dataset. In this paper, sensitive information is distorted with Min-Max, Z-Score, and Decimal Scaling privacy and accuracy normalization methods [7,17,18,19]. The technique proposed applies to the adult data set and the precision is compared with the decision tree algorithm J48 and the classification algorithm Naïve Bayes. Experimental results showed that data accuracy is maintained with the minimum loss of information by increasing the use of data. Two parameters evaluate the performance of the proposed technique: maintaining the high accuracy of the data mining application as well as the safeguarding of original data values.

II. RELATED WORK

Many literature works in the field of data mining in recent years have been undertaken to protect the confidentiality of sensitive information [8].Most of those works are divided into huge categories. They especially alter specific data mining algorithms. In the primary type, strategies, in general, modify the various data mining algorithms so that every one data mining operations take region while not having the knowledge of the facts miner's specific values. In the second one class type, strategies adjust the touchy values of information units in order that unique information values are protected. Several studies have been completed on data disturbance and statistics distortion techniques as follows:

Initially, the concept of transforming data set values into pairs of records was proposed by T. Dalenius et al. (1982). This technique is referred to as data exchange. They kept the data confidentiality and retained the lower frequency count. The experimental results have demonstrated that the data exchange technique preserves the original data values.

Liew et al (1985), defined a distribution of probabilities by the method of data distortion. This procedure works in three steps I Density Function Identification ii) Distorted series generation using density function iii) mapping of distorted series and original values.

Agrawal R. et al. (2000) have shown a new additive disturbance method for Classifier decision-making tree. Random noise such as gaussian distribution adds to each data element. This technique allows the reconstruction of original data from disturbed data.

Sweeney L. et al. (2002) defined a model based on k-anonymity that examined the problem of sharing sensitive data without revealing sensitive personal data. In this paper, techniques to delete and generalise sensitive information are used. This paper also discussed the re-identification attack.

Chen et al. (2005) proposed a new rotational data perturbation technique technique. This method maintained the accuracy of zero data loss for different classifiers. Experimental work has shown that rotational data perturbation technology improves the privacy quality without disregarding accuracy.

Weimin Ouyang et al. (2006) has established a multi-party, secure computational protocol that maintains the privacy of sensitive information using homomorphic encryption sequential pattern mining. By using the method proposed, each party can share the data with the trusted party by maintaining its confidentiality. This method can also be used to cluster and classify.

For perturbation, Saif et al. (2007) used a Non-negative Matrix Factorization(NMF) technique. The proposed method examined use of NMF truncated with Sparseness Constraints (Non-Negative Matrix Factorization). Experimental results show that Sparseness restrictions with non-negative matrix factorization is an efficient tool to protect privacy by mining data (PPDM).

The greedy technique used by Chieh-Ming Wu et al. (2009) is used to hide various sensitive values in a dataset. The difficulties in a normal rule hiding method are overcome. In this hiding place, confidential and sensitive values are concealed by the greedy method. The experimental results show that all sensitive rules are hidden without fake rules.

Peng et al. (2010) proposed a new technique based on combined data distortion techniques for data mining privacy conservation. The four schemes proposed, namely Single value decomposition (SVD), NMF (Non-Negative Matrix Factorization), DWT (Discrete Wavelet Transformation), were created to distort an original dataset sub-matrix in order to safeguard the privacy of original data values. Experimental results have shown that the proposed way of maintaining the privacy and data utility is very efficient.

Santosh Kumar Bhandare et al., (2011) illustrated a technique for data transformation by applying Z-Score Standardization to various data sets in real life. Experimental results have shown that an approach based on transformation is very efficient to maintain sensitive data.

G.Manikandan et al. (2012)" proposed a data transformation method for the preservation of sensitive attributes using the cluster method and obtained good accuracy results.

G. Manikandan et al. (2013) showed a new transformation-based technique using normalization. The proposed method obtained good results in data mining and high data accuracy.

G. Manikandan, C. Saranya#1 (2013), We analysed the use of normalization techniques to achieve privacy in this paper. We have compared these techniques and it can be concluded from the experimental results that Min-Max normalisation has a minimum error in classification.

Syed Md. Ahmad et al (2014), In this paper we use the minimum normalization approach to safeguard privacy during mining. We clean the original data with the minimum normalization before publication. We used k- means algorithms for experimental purposes and it is obvious from our results that our approach protects privacy and accuracy.

Patel Brijal H1., Ankur N. Shah2, et al., (2015) must develop algorithms to alter the originally collected data and to secure the misuse of the information to maintain privacy and private knowledge after mining. The topic reiterates a number of privacy safeguard technologies for data mining to protect the confidentiality of sensitive information and to obtain data clusters with minimum loss of information for multiple attributes in data sets.

MissAnjana Patel et.al.(2016) has defined various privacy measures and techniques for the protection of sensitive personal information. In this paper a novel method for protecting sensitive information with good accuracy results was described.

Kiran et al.,(2017) have illustrated various data mining privacy issues. Different data mining and data protection algorithms for data mining are very well explained in this paper.

Arif et al. (2020) suggested that a new scheme could secure privacy with increased capabilities for area extraction, which could significantly improve the threat to vehicle safety.

Song et al. (2019) proposed a method based on the encryption process in the latest study. The owner can use these public keys to encrypt the data. Data is passed to the user using his own encryption key.

Privacy systems have recently been developed by Abdelouahid et al. They examined strategies to ensure security of privacy. They analysed a new level of classification system, in which the latest privacy-safe learning methods demanded a rigorous degree of privacy attention.

III. Proposed Approach

The goal of normalization methods is to map the data to another scale. The literature contains different types of techniques for normalization [9]. In this paper we compared three standards including Min-Max Standardization, Z-Score Standardization and Decimal Scaling.

Min-Max Normalization:

Min-Max normalization is a well-used data transformation technique used in an individual dataset to maintain sensitive attributes. The value of the original data set is normalised using the min-max normalization function in a data set, taking minimum and maximum values into account. The Min-Max Standardization technology linearly transforms the original data set [10].

Each attribute in a record set is normalized to fit its values into the minimal consumer variety of 0.0 to 1.0. For Mapping a value,Vi of attribute A from the range of [minA,maxA] to [new_minA,new_maxA] is computed by following function

$$V_i' = \frac{\mathbf{Vi\text{-}min_A}}{\mathbf{max_A - min_A}}(\mathbf{new\_max_A\text{-}new\_min_A})\mathbf{+new\_min_A} \qquad (1)$$

Where V_i is the newly calculated cost in the user-particular range. By the usage of the Min-max normalization method (Eqn.1), the relationships between a few of the Original values are preserved.

Z-Score Normalization:

Z score normalisation, also known as Zero normalisation. The data is normalized here based on the mean and standard deviation. The formula is then

$$D' = \frac{\mathbf{D-Mean(P)}}{\boldsymbol{std(P)}}$$

Where Mean(P) = sum of the all attribute values of P

Std(P)=Standard deviation of all values of P

***Decimal Scale Normalization*:**

Normalization by decimal point-of-value movement of the attribute. The number of decimal points moved depends on the absolute maximum values of the attribute. The normalization form of decimal scale is,

$\boldsymbol{d'} = \frac{D}{10^m}$ (Where, m is the smallest integer that max (|d'|<1 (1)

Where d' is the newly calculated value in the user-specified range. By using the Decimal Scaling Normalization method (Eqn.1) the relationships among the Original values preserved.

3.2. Proposed Normalization Algorithm

Algorithm Procedure : Data transformation using normalization methods.

Input Values: Data Set D original, Sensitive Attributes S.

Intermediate _Output: Modified (Perturbed) data D'.

Output Values: *Results of* Classification R and R' for Datasets D and D' .

Step 1: Data set D with tuple size n and Sensitive attribute [S].
Step2: Applying normalization techniques on data with three popular data transformati techniques on the original information set

Step 2.1: Apply Min-Max Standardization on Sensitive Attributes [S].

$$V_i' = \frac{Vi\text{-}min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A$$

Step 2.2: Apply the decimal method of scaling to sensitive attributes [S].

$$D' = \frac{d}{10^m}$$ (where, m is the smallest integer that max (|d'|<1

Step 2.3: Apply Sensitive Z-Score Normalization method [S].

$$D' = \frac{d - Mean(P)}{std(P)}$$

Where Mean(P) = sum of the all attribute values of P
Std(P)=Standard deviation of all values of P
Step 3: You should perturb dataset D to make a D' dataset by making changes to the original dataset's attributes that have been deemed sensitive.
Step 4: Classify the original dataset D, which contains sensitive attributes, with the Naïve Bayes algorithm.
Step 5: Perturbed dataset D' with perturbed sensitive attributes S should be analysed using the Naïve Bayes classification algorithm.
Step 6: Using the procedure in Steps 4 and 5, analyse the results of analysing proposed Method's accuracy through Naïve Bayes Classification.
Step 7: Classify the original dataset D, which contains sensitive attributes, with the J48 algorithm.
Step 8: Perturbed dataset D' with perturbed sensitive attributes S should be analysed using the J48 algorithm.
Step 9: Using the procedure in Steps 7 and 8, analyse the results of analysing proposed Method's accuracy through J48 algorithm.

3.3. Proposed Normalization Architecture

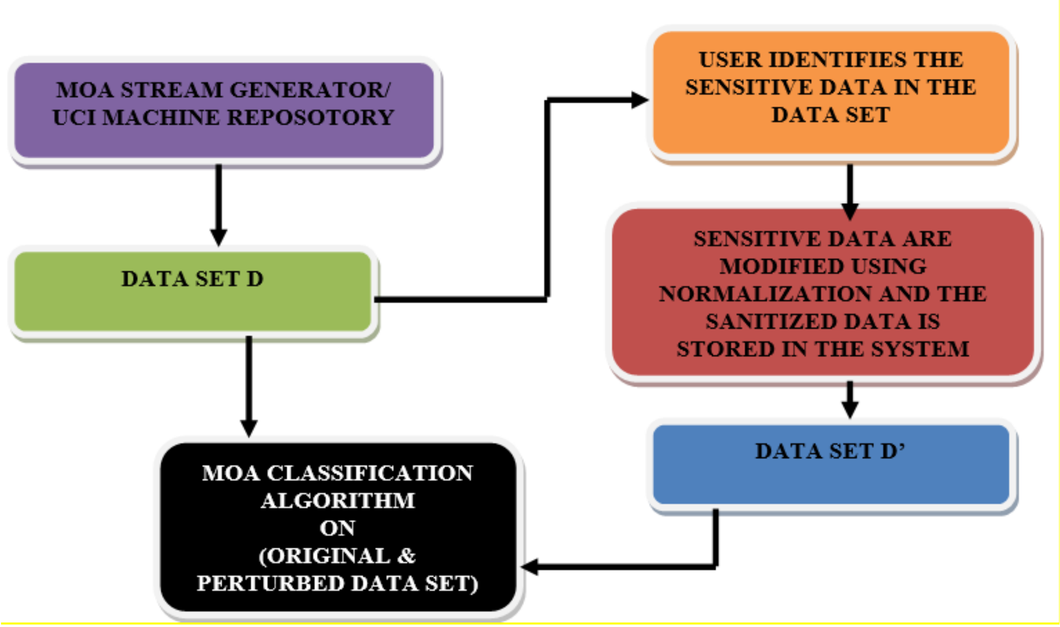
MOA STREAM GENERATOR/ UCI MACHINE REPOSOTORY
DATA SET D
USER IDENTIFIES THE SENSITIVE DATA IN THE DATA SET
SENSITIVE DATA ARE MODIFIED USING NORMALIZATION AND THE SANITIZED DATA IS STORED IN THE SYSTEM
DATA SET D'
MOA CLASSIFICATION ALGORITHM ON (ORIGINAL & PERTURBED DATA SET)

Fig. 1 Proposed Architecture of Normalization techniques

The purpose of the proposed method is to protect the sensitive characteristics of a dataset. This framework uses various techniques of data transformation, such as Min-Max Normalization for the protection of sensitive attributes and decimal scaling and Z-Score Normalization, to convert the original D data set into a disrupted d 'data set for high privacy [11].

Data mining is generally responsible for various data modification techniques for the protection of individuals' privacy in order to maintain sensitive values. The proposed methodology focuses mainly on two facts: Initially, it focuses on ensuring a level of privacy and, secondly, maximising data utility levels. The primary objective of each technology for data modification is to achieve greater privacy by ensuring maximum confidentiality and the use of sensitive values.

IV. Implementation

A. Data Set

Table 1: Description of (Adult Data) data set

Dataset	Description
Adult dataset	Attributes: 6 Total Instances: 48,842

Many datasets are available in the field of data mining to conduct research such as Bank Datasets, Cancer data sets, Cover data sets, retail datasets, Finance Datasets and Adult datasets. Our proposed methodology is analysed on adult data set in the UCI machine repository[12].

B. Experimental Results

Data Transformation techniques, using Min-Max Normalization, Scaling Score Normalization methods,[11] are used to implement the proposed methodology to transform original D data set values into disturbed data set D' values without loss of privacy.

Table 2 below shows a comparison of the NB (Naive Bayes Classification Algorithm) and J48 algorithms (Decision scaling analysis algorithm), the original values and the changed values of the education and its efficiency for the single column age, with three data transformation techniques (Min-Max Normalisation, Decimal Scaling Normalisation, Z-Score Standardization).

C. Evaluation Metrics

(i). Accuracy:

To understand the performance of the classifier, look at how well the classifier classifies and how it equates to the following equation [13].

$$\text{Classification Accuracy} = \frac{\text{Number of Tuples Correctly Classified}}{\text{Total Number of Tuples in the Dataset}} \quad \ldots\ldots\ldots\ldots \text{ (Eqn (4))}$$

(ii)RMSE: This proposed system calculates the average absolute error for each approach (MAE) and square root error (RMSE) (MAE). All differences are equally weighted, because the MAE is a linear score [14].

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n}(x-y)^2}{n}} \quad \ldots\ldots\ldots\ldots \quad \text{(Eqn (5))}$$

And MSE is calculated using the below formulae

$$MSE = 1/n \sum_{i=0}^{n} x - y \qquad \text{............} \qquad \text{(Eqn (6))}$$

Table 2: .Comparison of Naive Bayes for Attribute Using Data Transformation Techniques

NB Algoritthm	Age Column			
	Original classification values without Normalization	Proposed Naïve Bayes Value Using Min-Max	Proposed Naïve Bayes Value Using Decimal Scaling	Proposed Naïve Bayes Value Using Z-Score
CCI	83.25%	83.15%	83.30	83.09
ICI	16.74%	16.84%	16.69	16.90
KS	0.49	0.49	0.50	0.48
MAE	0.1746	0.176	0.1772	0.177
RMSE	0.3741	0.375	0.3722	0.3773
RAE	0.4795	0.4859	0.4867	0.4861
RRSE	0.8768	0.8798	0.8722	0.8843
TT (sec)	0.22	0.47	0.25	0.38

Table 2 shows the comparison between Naive Bays and Age attribute data processing techniques in the original dataset and disturbed datasets. In Table 2 you will also find several privacy measures such as RMSE, Kappa Statistics (KS), Relative Absolute Fehler (RAE), Relative Relative Squared Error (RRSE), Mean Absolute Error (MAE) etc.

Table 3.1:. NB classification results for age attributes Age Attribute with CCI and ICI Using data transformation techniques

NB (AGE Attribute)	Original Classification	Min-Max Normalization	Decimal Scaling Normalization	Z-Score Normalization
CCI	83.25%	83.15%	83.30%	83.09%
ICI	16.74%	16.84%	16.69%	16.90%

The original data set algorithm of Naive Bayes for the age attribute from Table 2.1 was 83.25% for correctly classified and incorrectly classified 16.74%. The accuracy of Naive Bayes algorithm for correctly classified Min-Max-standardizing instances in Perturbed data set was 83.15%, for decimal scaling 83.30%, for z-score normalization 83.09% and for perturbed datasets the incorrect classification was 16.84% for Min-Max-standardization. In decimal scaling, 16.69%, in Z-score normalisation, 16.90%. The confidentiality of the original data set is thus maintained in comparison with other methods with a lowest information loss with the decimal scaling method.

Table 3 : J48 classification results for age attributes Using data transformation techniques

J48 ALGORITHM COMPARISSION	Age Column			
	Original Classification Values without Normalization	Proposed J48 Value Using Min-Max	Proposed J48 Value Using Decimal Scaling	Proposed J48 Value Using Z-Score
CCI	86.07%	85.99%	86.00%	85.88%
ICI	13.92%	14.00%	13.99%	14.11%
KS	0.58	0.58	0.58	0.58
MAE	0.1956	0.1945	0.1937	0.1985
RMSE	0.3201	0.3211	0.3211	0.3238
RAE	0.5371	0.5343	0.5321	0.5452
RRSE	0.7502	0.7525	0.7525	0.7588
TT (sec)	6.46	9.02	13.25	12.4

Table 3 Displays the comparison between the Naive Bays in the original dataset and the disturbed Age Attribut data set with the use of the Data Transformation Techniques. Table 3 also shows a variety of technologies to measure privacy such as Root Means Square Error Value (RMSE), Kappa Statistics (KS), and the amount of time needed to complete the task.

Table 3.1 .Classification results of J48 for Age Attribute with CCI and ICI Using Data Transformation Techniques

J48 (AGE Attribute)	Original Classification	Min-Max Normalization	Decimal Scaling Normalization	Z-Score Normalization
CCI	86.07%	85.99%	86.00%	85.88%
ICI	13.92%	14.00%	13.99%	14.17%

In Table 3.1, the accuracy of the J48 algorithm with the original data set for age attribute was 86.07% for the correctly classified instance, and 13.92%. J48 algorithm for disturbed datasets was correctly classified with min-max normalisation at 85.99 percent, with 86.00 percent decimal normalization, with min-max normalization at 85.88 percent and with disturbed data set 14.00 percent incorrectly classified. 13.99 percent for Decimal Scaling, 14.17 percent for Z-Score Standardization. The confidentiality of the original data set is thus maintained in comparison with other methods with a lowest information loss with the decimal scaling method.

Table 4 .Comparison of NB for Education Attribute Using Data Transformation Techniques

NAÏVE BAYES COMPARISSION	Education Column			
	Original Classification Values Without Normalization	Proposed Naïve Bayes Value Using Min-Max	Proposed Naïve Bayes Value Using Decimal Scaling	Proposed Naïve Bayes Value Using Z-Score
CCI	83.25%	83.04%	83.02%	80.72%
ICI	16.74%	16.95%	16.97%	19.27%
KS	0.49	0.48	0.47	0.36
MAE	0.1746	0.1772	0.1774	0.1934
RMSE	0.3741	0.3755	0.374	0.3883
RAE	0.4795	0.4868	0.4871	0.5311
RRSE	0.8768	0.8800	0.8767	0.9102
TT (sec)	0.22	0.44	0.38	0.5

Table 4 shows the comparison of Naive Bayes with the original dataset and disturbed data set on data transformation techniques. Table four Table 4 also displays different technical data protection measures such as Root Means Square Error (RMSE), Kappa Statistics (KS), Relative Absolute Error (RAE) Root Squared Error (RRSE), Middle Absolute Error, etc.

Table 4.1 .Classification results of NB for Education Attribute with CCI and ICI Using Data Transformation Techniques

NB (Education)	Original Classification	Min-Max Normalization	Decimal Scaling Normalization	Z-Score Normalization
CCI	**83.25%**	**83.04%**	**83.02%**	**80.72%**
CI	**16.74%**	**16.95%**	**16.97%**	**19.27%**

Table 4.1 indicates 83.25% of accuracy of the Naive Bayes algorithm in the original Education Dataset attributes for correctively classified instances and 16.74% of incorrectly classified instances. At same time, with the correctly classified min-max standard instances, the accuracy of Naive Bayes Perturbed Dataset was 83.04%, the decimal scale was 83.02%, the z-score normalisation was 80.72% and the incorrect min-max normalization data set was 16.97 percent, the Z-score normalisation figure was 19.27%. The privacy of the original data set is therefore maintained by standardising Min-max compared to other methods with minimal information loss.

Table 5 .Comparison of J48 for Education Attribute Using Data Transformation Techniques

J48 ALGORITHM COMPARISSION	Education Column			
	Original Classification Values Without Normalization	Proposed J48 Value Using Min-Max	Proposed J48Valu Using Decimal Scaling	Proposed J48Value Using Z-Score
CCI	86.07%	85.92%	86.21	85.82%
ICI	13.92%	14.07%	13.78	14.17%
KS	0.58	0.58	0.58	0.58
MAE	0.1956	0.1957	0.2003	0.3233
RMSE	0.3201	0.3211	0.3232	0.3211
RAE	0.5371	0.5374	0.5501	0.5570
RRSE	0.7502	0.7526	0.7575	0.7578
TT (sec)	6.46	11.72	11.94	10.09

Table 5 Shows the J48 comparison of data transformation techniques in the original data set and the disrupted data set of education attributes. The Tables 5 display various privacy measurement values such as Kappa statistics (KS), Root Media Square Error (RMSE), Mean Absolute Error (MAE), Relative Absolute Error (RAE), etc.

Table 5.1 .Classification results of NB for Education Attribute with CCI And ICI Using Data Transformation Techniques

J48 (Education)	Original Classification	Min-Max Normalization	Decimal Scaling Normalization	Z-Score Normalization
CCI	**86.07%**	**85.92%**	**86.21**	**85.82%**
ICI	**13.92%**	**14.07%**	**13.78**	**14.17%**

Table 5.1 shows that the accuracy of the algorithm J48 on the original data set of the education attribute of 86.07% was wrongly classified and 13.92%. At the same time, accuracy of J48 on Perturbed Dataset was 14.07% with Min-Max Normalization in correctly classified instances, 86.21% for decimal scales, 85.82%, and inadequately classified with disturbed data set was 85.92% Min-Max Normalization For decimal scaling, 13.78%, for Z-score normalization, 14.17%. The privacy of the original data set is therefore maintained by standardising Min-max compared to other methods with minimal information loss.

Table 6: Comparison of Naive Bayes for Attribute and Education Using Data Transformation Techniques

Adult Data Set	Original Classification Values Without Normalization	Proposed Naïve Bayes Value Using Min-Max	Proposed Naïve Bayes Value Using Decimal Scaling	Proposed Naïve Bayes Value Using Z-Score
AGE	**83.25%**	**83.15%**	**83.30%**	**83.09%**
EDUCATION	**83.25%**	**83.04%**	**83.02%**	**80.72%**

Table 6 shows the accuracy of the Naïve Bayes algorithm on original data sets and disturbed educational and age-related data together with results of data transformation in the age and educational attributes. The results of the original age attribute Without transformation technology, after application of Min-Max Normalisation, classification results are 83.25%, 83.15%. The precision after use is 83.30 percent and the results of the Z-score classification are 83.09 percent The original results of education classification Unless transformation technology is employed, after Min-Max normalization, the result is 83,04%, after the application of Decimal Scaling, precision is 83,02% and the result of Z-Score rating is 80,72%. Finally, the decimal scaling method gives better results of the normalisation procedure for the age and min-max, providing the results of classification of educational attributes. The minimum information loss has been maintained in addition to the accuracy of the proposed framework. With minimum information loss, the privacy of the original data can be maintained.

Table 7. Comparison of J48 for Age and Education Attribute Using Data Transformation Techniques

Adult Data Set	Original Classification Values Without Normalization	Proposed J48 Value Using Min-Max	Proposed J48Value Using Decimal Scaling	Proposed J48Value Using Z-Score
AGE	86.07%	85.99%	86.00%	85.88%
EDUCATION	86.07%	85.92%	86.21%	85.82%

The accuracy of J48 data sets and the disturbed data sets of education and age, along with results from data transformation for ages and attributes of education, are shown in Table 7. The results of the original age attribute The classification result after Min Max Normalization is 85,99 percent, the precision after the Decimal Scaling is 86,00 and the result of Z-Score Classification is 85,88 percent without application of the transformation technology. The original results of education classification Without the application of the technique, the classification result after Min-Max Normalization application is 85.92 per cent and precision is 86.21 per cent following the application of Decimal Scaling. Ultimately, the decimal scaling method results in better results for both age and education attributes. A minimum information loss was also maintained, in addition to the accuracy of the proposed framework. With minimum information loss, the privacy of the original data can be maintained.

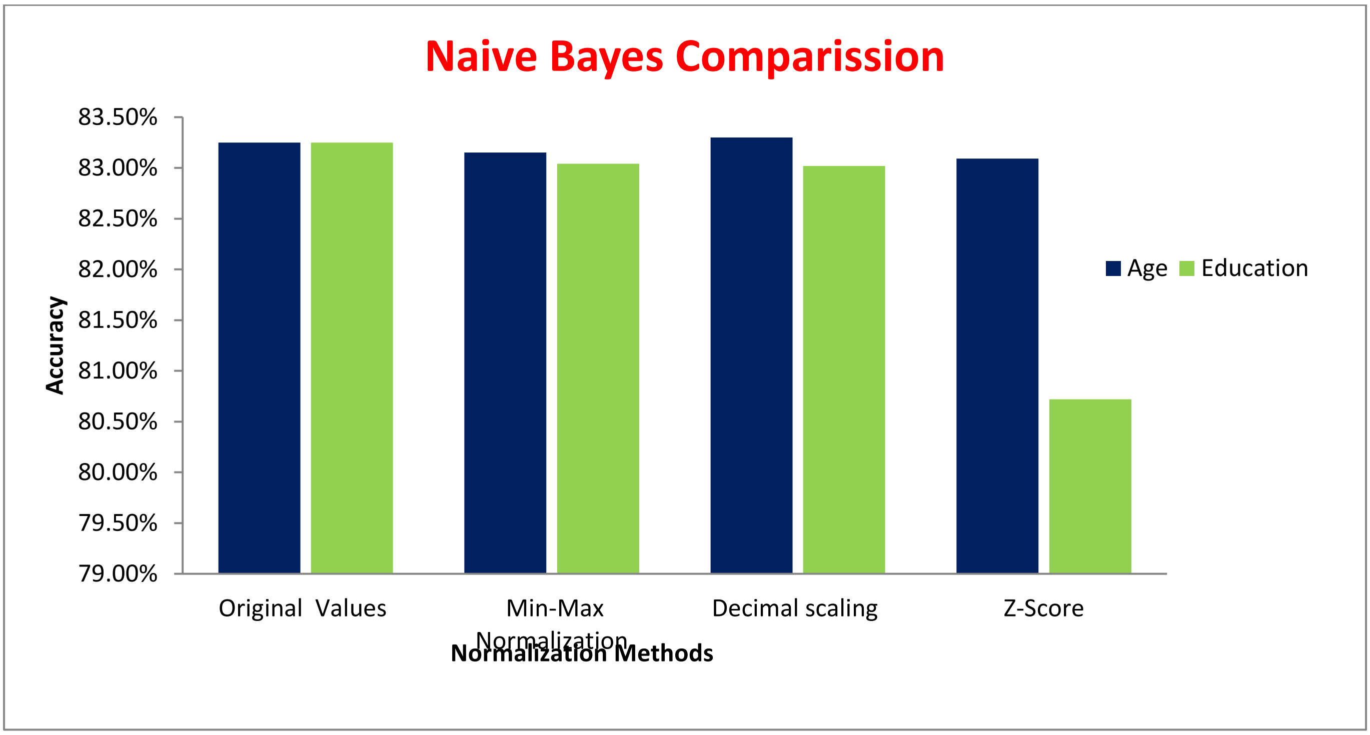

Fig. 2. Illustrates the Comparison of Existing and Proposed with J48 Algorithm

The accuracy of age and education attribute results can be compared with the Adult Dataset Classification Algorithm in Naïve Bayes and data set sets of perturbed data. For all three data transformation techniques, original data set values and minimum information loss are maintained.

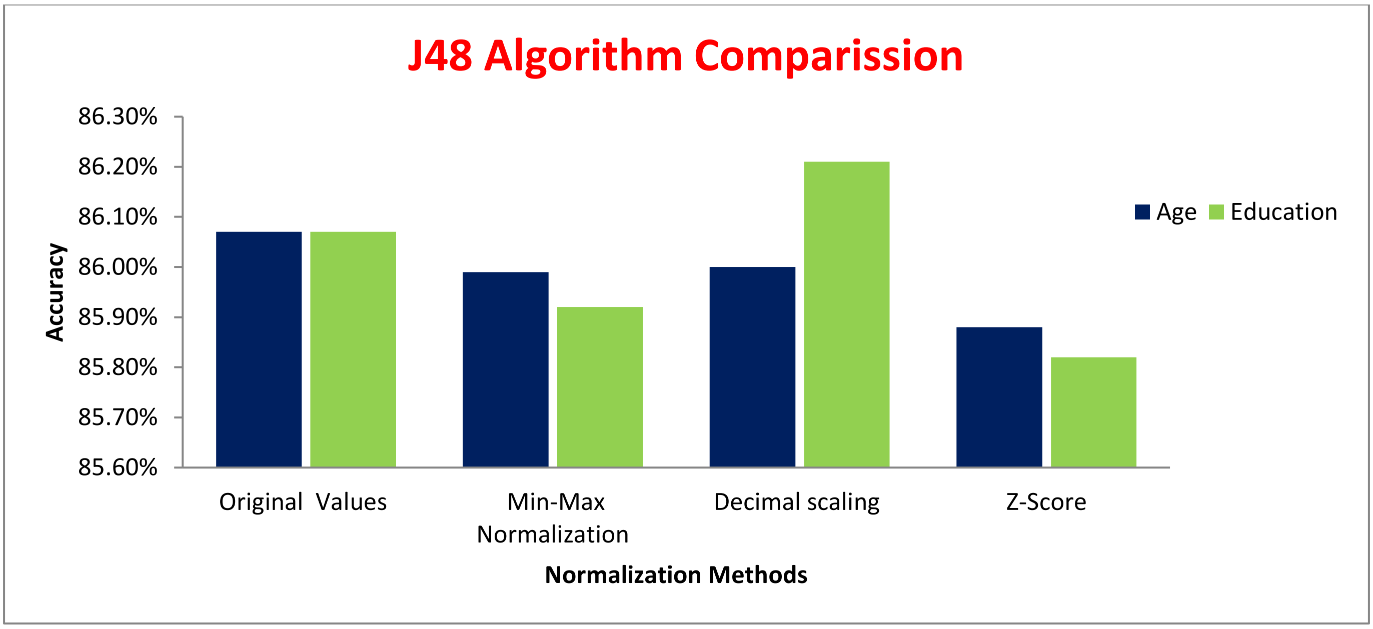

Fig. 3. Illustrates the Comparison of Existing and Proposed with J48 Algorithm

In Above Figure 3, results of age and training accuracy are compared with J48 adult decision tree algorithms and Perturbed data sets. For all three data transformation techniques, original data set values and minimum information loss are maintained.

V. CONCLUSION

This paper gives a developing way to deal with building a practical anticipations illustrate for flexible source provisioning in the reasoning motivate powerful and practical source management, reservation and opportunity company for brilliant web centered company programs where explosiveness and responsiveness are basically crucial. All through the evaluation, we have evaluated a few popular machine learning computations, specifically varying moving window measure with a view to giving exact expecting early. We have shown our suggested anticipations systems with regards to the data-set obtained by using TPC-W, is a complete structure, which is resolved for web centered company related advantages. We additionally provided evaluation dimensions to accepting the truth of the suggested anticipations strategies and looked at the performance of the predetermined strategies using these dimensions. The category perfection of applied methods is strengthening and show frequent adequacy of category procedure i.e. Sensory Network shows and determine source use in the reasoning. Further enhancement of our suggested approach is to improve source provisioning with specific to enhanced cost performance in service utilized for reasoning computing.

VI. REFERENCES

[1]. M. Chen, J. Han, and P. Yu, "Data mining: An Overview from a database Prospective", IEEE Trans. on Knowledge and Data Engineering, vol. 8, no. 6, pp. 866-883, Dec. 1996.

[2]. Ajmeera Kiran , D. Vasumathi, “Optimal Privacy Preserving Technique over Big Data Analytics using Oppositional Fruit fly Algorithm” Recent Patents on Computer Science , Vol. 11, Issue: 0, DOI :10.2174/2213275911666181119113913, pp: 1-12 , 2018.

[3]. C. P. Chen and C. Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data”, Information Sciences, Vol. 275, pp. 314-347, 2014.

[4]. Atallah, M., Elmagarmid, A., Ibrahim, M., Bertino, E., Verykios.V:Disclosure limitation of sensitive rules, Workshop on Knowledge and Data Engineering Exchange, 1999.

[5]. K.Liu, H Kargupta, and J.Ryan,” Random projection–based multiplicative data perturbation for privacy preserving distributed data mining .” IEEE Transaction on knowledge and Data Engg.Jan,pp.92-106,2006.

[6]. Kiran Ajmeera, Vasumathi D, “A Comprehensive Survey on Privacy Preservation Algorithms in Data Mining” Proceedings of the 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC-2017), 978-1-5090-6621-6/17/$31.00 , 2017.

[7]. UCI Machine Learning Repository http://archive.ics.uci.edu/ml/datasets.html.

[8]. J.Nageswara Rao, M.Ramesh,” A Review on Data Mining & Big Data, Machine Learning Techniques”, International Journal of Recent Technology and Engineering (IJRTE) , ISSN: 2277-3878, Volume-7 Issue-6S2, April 2019.

[9]. D. Veeraiah and J. N.Rao, "An Efficient Data Duplication System based on Hadoop Distributed File System,"2020 International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 197-200, doi: 10.1109/ICICT48043.2020.9112567.

[10]. J. N. Rao, A. C. Singh, “A novel encryption system using layered cellular automata,” International Journal of Engineering Research and Applications, vol. 2, no. 6, pp. 912–917, 2012.

[11]. J.Nageswara Rao, Bhupal Naik, G Sai Lakshmi, V Ramakrishna Sajja, D Venkatesulu, “Driver’s Seat Belt Detection Using CNN” Turkish Journal of Computer and Mathematics Education Vol.12 No.5 (2021), 776-785 3.

[12]. J.N Rao, Dr.Rambabu Busi, Dr. G Rajendra Kumar, U. Surya Kameswari, “ Content image Retrieval Based on using open Computer Vision and Deep Learning Techniques “International Journal of Advanced Science and Technology,Volume29Issue03Pages5926 - 593 92020)

Cite this article as :

Surendra Kumar Reddy Koduru, "Linear Regression based Aggressive Resource Provisioning for Cloud Computing", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 5, pp. 144-151, September-October 2022. Available at doi : https://doi.org/10.32628/CSEIT1228529

Journal URL : https://ijsrcseit.com/CSEIT1228529