



A Data Centric Privacy Preserved Mining Model for Business Intelligence Applications

¹Prof. Dr. P.K. Srimani, F.N.A.Sc. ²Prof. Rajasekharaiah K.M.

¹Former Chairman, Dept. of Computer Science & Maths, Bangalore University
Director, R & D, Bangalore University, Bangalore,
India.

² Professor & HOD, Department of Computer Science and Engineering,
JnanaVikas Institute of Technology, Bangalore Mysore High Way, Bidadi, Bangalore,
Visvesvaraya Technological University (VTU), Belgaum, Karnataka
India

Abstract: In present day competitive scenario, the techniques such as data warehouse and on-line analytical process (OLAP) have become a very significant approach for decision support in data centric applications and industries. In fact the decision support mechanism puts certain moderately varied needs on database technology as compared to OLAP based applications. Data centric decision support schemes (DSS) like data warehouse might play a significant role in extracting details from various areas and for standardizing data throughout the organization to achieve a singular way of detail presentation. Such efficient data presentation facilitates information for decision making in business intelligence (BI) applications in various industrial services. In order to enhance the effectiveness and robust computation in BI applications, the optimization in data mining and its processing is must. On the other hand, being in a multiuser scenario, the security of data on warehouse is also a critical issue, which is not explored till date. In this paper a data centric and service oriented privacy preserved model for BI applications has been proposed. The optimization in data mining has been accomplished by means of C5.0 classification algorithm and comparative study has been done with C4.5 algorithm. The implementation of enhanced C5.0 algorithm with BI model would provide much better performance with privacy preservation facility for Business Intelligence applications.

Keywords: Business intelligence, Data warehouse, OLAP, data mining, privacy preservation, C5.0, rule generation

I. INTRODUCTION

Numerous organizations and business houses need certain system model that can effectively maintain huge data collections and can perform its processing for further analysis. The information retrieved from its local databases might not be efficient to accomplish the expected results for getting certain decision making solution. Therefore, in order to accomplish such objectives, it is required to have a framework that can collect huge datasets from diverse areas and can perform mining to achieve the expected help in service oriented organizational decision making process. In present day scenario the intricacies existing in business environment needs certain potential facilities that could make organizations swift and proactive for management oriented optimistic decision-making. Industry needs higher production and it can be facilitated by an effective resource planning that can retain its enterprise survival in competitive markets. In recent era each complicated and long-term information approach is suffered with the issues related to its business rules [1]. Then while, majority of business rules management approaches are emphasized on data collection, processing and its ultimate detail reporting. The detailed reporting can be a significant achievement for optimum decision making rather than to have good knowledge only [2]. The service oriented industries need highly intelligent production and resource management schemes for predicting the organizational performance and for providing short-term arrangement and control substitutes. Data warehouse is nothing else but a “subject-oriented, integrated, time -varying, non-volatile collection of data that can be employed predominantly for effective organizational decision making processes and business intelligence (BI)” [3].

With the help of BI approaches the organizations can effectively integrate certain potential analytical tools for standardized data reporting, monitoring and analysis while taking into account of numerous metrics in certain service-oriented architecture [4], [5] which is must for any kind of management and strategic planning. The data analysis and its revelation might detect evolutionary data ignored in the past. Ultimately employing strategic measures an alignment can be done with certain tactical schemes while incorporating stringent objectives to be accomplished every moment, permitting permission on simplified and highly robust approach

for correction in reports divergences. Such feasibility in historical analysis can be accomplished in case the organization possesses certain data bases or data from internal as well as external resources in data warehouse, facilitating investigative data analyses while implementing Business Intelligence (BI) approaches [6]. This integration of data warehouse and OLAP frameworks with BI tools can facilitate an optimistic strategic decisions support system which is significant in the course of exploratory schemes with huge data analyses with unknown patterns. **Figure 1** illustrates management framework for performance which exhibits that data mining enhancement can play a very vital role in overall Business intelligence optimization.

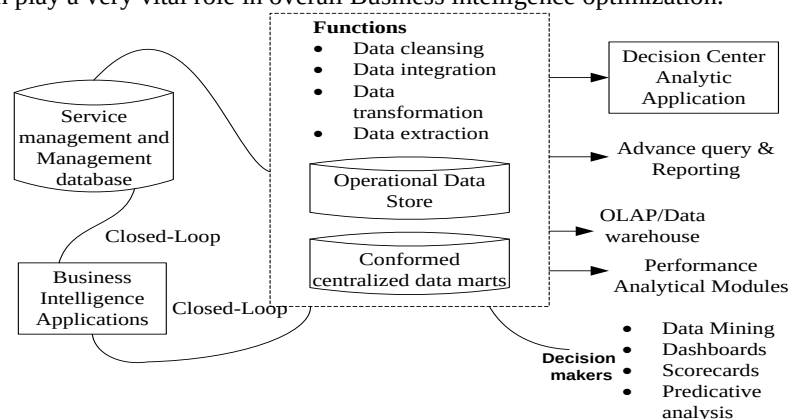


Figure 1: Performance management framework

For strengthening the performance of data warehouse and BI techniques, it is always advocated to enhance the data mining scheme. The effective data processing and information reporting can be accomplished only with better mining and classification approaches. Meanwhile, data security is a vital issue in data warehouse applications and business application tool development. The privacy preservation of data in this entire data mining, processing and analysis process might play a significant role [7][8], therefore in this paper, a highly robust privacy preservation enriched data mining algorithm has been developed for BI applications. Researchers always focus on employing certain optimum decision tree classification algorithm for BI frameworks. The C5.0 algorithm facilitates optimum data mining for further data analysis. For BI process and data warehouse applications C5.0 algorithm has been employed for performing analysis of data sets associated with various classified classes or parties. Since for classification accuracy and optimized rule generation of C5.0 can perform better as compared to C4.5 [9][10][11][12], in this paper, C5.0 data mining algorithm has been implemented for performing rules generation and data classification. The optimized data mining approach can be implemented with any data warehouse or online analytical process (OLAP) frameworks and business intelligence tools. The optimized mining process with enriched privacy preservation would provide better results for data centric and service oriented BI utilities. Other section of this paper is arranged in this way: Section II presents related works; in Section III proposed framework has been discussed. In Section IV the business intelligence oriented data analysis using C5.0 algorithm has been presented. The integration of OLAP with proposed mining module has been presented in Section V. Section VI present the results obtained which is followed by conclusion in Section VII.

II. RELETED WORKS

A number of systems have been proposed for enhancements in BI or decision support and mining modules are found as a potential solution for it [13]. Work [14] advocated a hybrid approach of data mining and analytical scheme for decision support in BI. They also developed a unified presentation for BI architecture using stream mining concept. In [15] a BI driven data mining (BidDM) model was proposed for decision support in e-commerce, comprising four layered framework and mining model. Chang et al in [16] developed a BI system with web information while performing data integration with mining model. In data integration they employed Oracle database with Cognos BI tool and in mining module they implemented multimedia data analysis using Monago database with Pentaho BI tool and performed comparative analysis, which states that data mining can exhibit better for BI. In [17] a BI model was developed for bankruptcy prediction and its prediction using qualitative as well as quantitative factors using Genetic algorithm [18]. A similar work was done in [19] with a mining module for rule generation for fraud patterns. Precise decision making by means of correct and reliable data in data warehouse was explored in [20] [21]. In [21] data sets from UCI machine learning repository was processed for data clustering and its visualization for significant decision making. In data mining for security reasons, in [22] a k-anonymity based privacy preservation scheme was developed while in [23] system was enhanced by classification and weighted attributes. In [24][26] a privacy preservation model for distributed decision tree mining module was developed on the basis of homomorphic encryption scheme. For OLAP application work [25] presented an approach of privacy preservation for OLAP multi-dimensional data cubes that enhanced the visualization of data sets in various perspectives in BI utilities. This is the matter of fact that

these works have come up with certain enhanced outcomes, but considering an optimized solution for BI application and data warehouse, still there exist gap for further enhancements, such as highly efficient rule set generation, accuracy in classification and swift processing for huge data sets. Majority of existing systems employ ID3 or C4.5 algorithm for mining but it can further be enhanced by means of C5.0 algorithm which can provide better results for privacy preserved data mining for business intelligence (BI) applications. In this paper, we have considered C5.0 algorithm to accomplish this goal.

III. PROPOSED FRAMEWORK

An optimized framework might be an effective approach to perform validation for various needs of certain project, in the form of a scheme to detect any irregularity and discrepancy against the specific needs allied with the project. Considering the need of precise, privacy preserved and accurate data processing system for BI in this paper we intend to develop a data centric and service oriented, privacy preserved mining system that can provide optimally classified and accurate data classification. Such enhancements in mining and classification would strengthen the effectiveness and performance of BI model. The optimized mining approach would be efficient for its implementation in OLAP or data warehouse applications for BI tools. In this paper we have proposed C5.0 classification algorithm enriched with data privacy preservation approach. Privacy preservation scheme with C5.0 would make system highly efficient in terms of data classification, accuracy and anonymity in entire data warehouse. The mining of huge datasets from various resources in BI function would provide optimistic decision making while exhibiting data extraction, its processing and ultimate visualization of data retrieved from huge data from unknown patterns. The proposed framework has been illustrated in **Figure 2**.

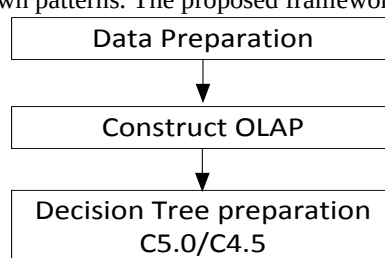


Figure 2: Proposed BI Framework

The mining enriched model can be integrated with data warehouse of OLAP to yield better results and decision supports in BI applications.

A. System Model

In data warehouse or OLAP frameworks, the collected datasets might encompass various kinds of integration anomalies. On the other hand being in multiuser scenario, there might be certain possibilities of data misplacement or even breaching. Therefore the datasets available are required to be cleaned and preserved for its privacy or anonymity. Once the data has been preserved for its security then it is succeeded by construction of OLAP multidimensional cubes which is followed by decision tree preparation using optimized C5.0 algorithm. The data mining can be exhibited on any dimension of the cube generated in second step. Once the model is prepared, it is then stored in OLAP cube where each dimension of cube represents the rules associated with certain node in the decision tree mining model. A sequential flow for the developed system has been presented in **Figure 3** as follows:

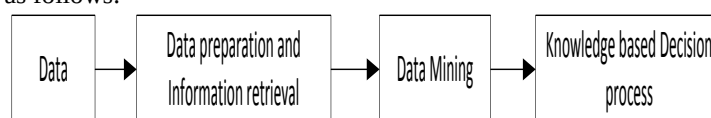


Figure 3: BI functional flow chart

A.1 Data Preparation:

The overall process of data preparation is classified into three sequential phases. These are, selection of data, data cleaning and formatting of data. In first step, the data can be selected on the basis of certain criteria, such as its data completeness and accuracy. Similarly, data selection also depends on its technical constraints like maximum data limits or data type, which is in fact associated with mining tools that we have planed earlier to operate with. The second step performs data cleaning that encompasses process such as data normalization and either the decimal scaling into range (0,1) or standard deviation normalization. Data cleaning can also have processes such as data smoothing, rectification for missing parameters, reduction in data volume in case of ineffective or even imperfect. In our work we have employed the discretization of numeric attributes for exhibiting data smoothing. For rectifying the issue of missing values, we have employed a common way called replacement of missing value with a single global constant. The data dimension reduction has been done by analyzing predictive potential of the specific data to be employed for decision making. This is accomplished

using schemes of attribute selection with mean and standard variation, principal component analysis (PCA) or feature merging using linear transform. In this work we have considered attribute selection process while taking into account of mean and variances. Once the data has been cleaned then it is followed by new data construction that performs constructive functions such as deviation of certain attribute from two or more attributes in the data sets, generating new data sets, its transformation (data normalization and smoothing) and privacy preserved data aggregation from different tables. In this work, the data formatting is achieved by syntactic modification to the specific data sets which do not change its significances. It encompasses re-arrangement of data attributes, variation in constraints as per tools, removal of any unwanted entities like comma, special characters, tabs etc. It can also do certain trimming for higher data size.

A.2 OLAP Cube Construction

In business intelligence (BI) applications the significance of OLAP is to exhibit materialization of certain vital but expensive computations which is inquired frequently. These expensive computations encompass processes such as aggregate function, counting, summation, averaging, max estimation etc and stores the processed datasets on certain multidimensional databases (Cubes) for supporting decision making in BI applications. Here it must be noted that “cube” states for data base. In OLAP formation the data aggregation operation might be a function that can be pre-estimated as per classification accomplished by various datasets and attributes. The data values in individual attribute could be categorized into certain lattice architecture of hierarchies. The process such as specialization and generalization can be exhibited on multi-dimensional databases by employing paradigms called “roll-up” or “roll-down”. The roll-up process minimizes the number of data dimensions and performs generalization on attributes values to the higher level approaches. On the other hand, the drill-down approach exhibits the reverse of the roll-up process. As, a number of aggregate functions might be required to be estimated iteratively for data analysis and BI applications, the processed data storage (pre-estimated results) can be done in multi-dimensional cubes that would ensure high speed computation and thus the response time would be enhanced. This also enhances flexibilities in viewing data originating from various angles and with numerous level of abstraction.

A.3 Decision Tree Construction

In this paper, we have implemented C5.0 algorithm for data classification and decision tree generation that can be employed for achieving most precise and accurate data information. In this work we have emphasized on the implementation of C5.0 algorithm and a performance oriented comparison has been done with C4.5 decision tree classification algorithm.

(a) C5.0 Data Mining Algorithm

The decision tree classification algorithm C5.0 is the successor of C4.5 algorithm. The C5.0 algorithm exhibits better efficiency and accuracy as compared to its descendant C4.5 algorithms, CART or ID3 algorithm. OLAP framework considered C5.0 algorithm to provide optimum classification and rule set generation which is must for BI tools. In this proposed framework, C5.0 algorithm has been used for creating rule sets which is further enhanced for classifying data sets from different sources or attributes of organization. The information available with attributes, called as Entropy, plays vital role in justifying effectiveness of decision tree classification learning process. For a random variable n , as per information theory, entropy can be defined as $(n) = \sum_n v(e) \log_2 \frac{1}{v(e)}$, Where e states attributes and $v(e)$ signifies for the value of the class attributes. Meanwhile, for another attribute the conditional entropy is given by $H(n/m) = \sum_{n,m} v(n,m) \log_2 \frac{1}{v(n,m)}$. For a better system model, the conditional entropy $H(n/m)$ is needed to be made lower than entropy for x $H(n/m) \leq H(n)$. In uncertain scenario this type of minimization is stated as mutual information existing between attributes given by $I(n; m) = H(n) - H(n/m)$. Primarily, for making each data partition the attribute m_i is optimized that could give maximum information about other attribute n and thus the overall information gain $I(n; m)$ is enhanced. As this process considers the employment of a parallel approach to assist m_i , resulting into multiple results, the classification algorithm C5.0 enhances the relation $\frac{I(n, m_i)}{H(m_i)}$ that provides *Gain ratio*. Additionally, to avoid the assortment of lower entropy by any attribute, *Gain ratio* factor can be optimized by an adaptive weighting process. The dynamic weight update is done by incrementing $H(m_i)$. A predominant problem in decision tree classification, rule set/rule generation schemes is the compliance of training data sets which is in general stated as overfitting. For eliminating this kind of limitation, the proposed C5.0 algorithm advocates a scheme stated as *Pruning*. Here, the reason behind using C5.0 algorithm with data warehouse applications are its characteristics; adaptive boosting, pruning and enhanced rule set generation that can optimize business intelligence tools for precise decision making. A brief of these features have been given in next section.

(b) Adaptive Boosting

The reason behind the implementation of adaptive boosting in C5.0 is the generation of a number of classifiers on the data existing in different cubes (consider individual party). In case of any unique data occurrence for

classification, the predicted data class is adaptively weighted with the help of independently trained data classifiers. C5.0 generates multiple data classifiers by inducing a single data classifier. The induced classifier is further re-trained on the data samples considered for forming the initial classifier. Final data sets are generated using second classifier. The significant algorithmic approaches behind the realization of adaptive boosting are:

- **Step-1:** Choose d data samples from existing datasets, where each datasets provides a likelihood of $1/P$ to train certain data classifier.
- **Step-2:** Classification of data sets with the trained classifier.
- Replace the data present by multiplying the likelihood of the imperfectly-classified datasets by certain adaptive weight W .
- **Step-3:** Go on with earlier steps k times with the generated likelihoods.
- **Step-4:** Blend the classifier C by introducing a weight factor $\log(CW)$ on independently trained data classifier.

(c) **Pruning**

The proposed C5.0 algorithm exhibits two consecutive steps for generating decision trees. Initially it constructs a data classifier for available data sets and in ascending phase it performs pruning on the data classifiers formed in initial phase. This is done only to prevent the probability of over-fitting on available data in individual datasets or party (cube). In this entire process two significant influent alternatives can be considered to pruning. The initial phase functions for characterizing the degree in which the decision tree generated is adaptable with training datasets. To accomplish it, this approach finds out minimum number of datasets which might have minimum two branches at any of the generated nodes in decision tree. Here, the predominant goal is to avoid data over-fitting by ending training. Another option for pruning is that its security features which do influence this algorithm states for the process of post pruning on the generated decision trees and generated rule sets. On the other hand, pruning process is also performed by removing some generated decision trees that do not have more significance due to higher error rate. The post pruning is exhibited by substituting a branch of classified tree by a leaf in the case when the predicted errors for the last step is lower than one of the branch in decision tree.

(d) **Rule Sets Generation**

One of the predominant functions of C5.0 algorithm is the conversion of decision trees into optimal rule sets. Here in proposed mining model the algorithm has been optimized in such a way so that it can generate rule sets for classifying data. Our proposed system model has exhibited optimized results in terms of flexible rule generation and the generated rule sets are easy to understand for decision making applications in BI models. The rule generation plays very significant role in optimizing performance of Data warehouses, OLAP and the ultimate BI tools. This is the matter of fact, that in data warehouse, there are huge data sets and attributes which might be operated by multiple users or operational attributes, therefore security or precisely anonymity of data is an important need. Considering this requirement of privacy preservation in mining or data warehouse systems, in this paper we have developed a privacy preservation approach that ensures authenticity and anonymity of data elements across data warehouse infrastructure.

IV. **BI ORIENTED DATA ANALYSIS with C5.0 ALGORITHM**

Considering the significance of a robust decision tree classification system with privacy preserved security features for data mining in data warehouse and business intelligence applications, in this paper the C5.0 algorithm [26] has been enhanced with privacy preservation. The efficiency of C5.0 can be visualized in terms of higher accuracy in classification, rule generation and swift processing with huge data sets. These all aforementioned factors are must for effective business intelligence (BI) or decision support applications. In this paper, C5.0 algorithm is optimized for its function for data analysis with raw as well as unclassified data available with every data set or cubes (in terms of OLAP). Once the rules generated by individual classifier are collected then these are combined together to exhibit higher accuracy and swift data processing. Let, an entity Sr_d represents a set of row or unclassified data available with datasets Cub where $Cub_d \in Cub$, and for analysis process the mingled rules generated with d collected classification rules from each party D_{Set} then the combined rule set generated is expressed by $D_d^{C5.0} = C_{RrmRG}^{C5.0}(D_{Set}, d)$, where D_{Set} represents local rule set formed on the basis of the data Sr_d and $C_{RrmRG}^{C5.0}$ that refers C5.0 enriched rule set generation function. $D_d^{P5.0} = C_{RrmRG}^{C5.0}(\{Dr_1 \cup Dr_2 \cup Dr_3 \dots \cup Dr_d\}, d)$; Here Dr represents rule set formed. $D_d^{C5.0} = C_{RrmRG}^{C5.0}(\{\{Rr_1, Rr_{21} \dots, Rr_{1mCub1}\} \cup \{Rr_{12}, Rr_{22} \dots, Rr_{1mCub2}\} \cup \dots \cup \{Rr_{1n}, Rr_{2n} \dots, Rr_{1mCubd}\}\}, d)$; Where Rr represents the maximum number of rules generated by certain cube or party. Now, observing these expressions it can be realized that the final combined rule sets generated by our proposed C5.0 algorithm ($D_d^{C5.0}$) may encompass every rule generated by Cub_1 to Cub_d . This would strengthen maximum utilization of data sets available for decision making. The utmost feasible number of rules generated by a party Cub_d is referred by Rr_{1mCubd} . The $C_{RrmRG}^{C5.0}$ Represents for the function of overall combined rule sets generation. The expression $fc_d^{C5.0} = C_{cls}^{C5.0}(D_d^{P5.0}, Sr_d)$ is the function for exhibiting classification by means of C5.0 algorithm. Here, $C_{cls}^{C5.0}$ depicts a function which considers two factors; classification rule $D_d^{P5.0}$

and input row data available Sr_d in n datasets or cubes in OLAP. The pseudo code for data analysis is given in **Figure 4** as below:

Business Intelligence oriented data mining and analysis using C5.0 Algorithm Input: Combined rule sets generated D_{Set} Output: Analyzed and mined data results $fc_d^{c5.0}$ 1. Initializing combined rule set $D_d^{c5.0} = \emptyset$ 2. Initialize $D_{pdt} = 0$ 3. For each $Dr_d \in D_{pool}$ 4. For each $Rr_{1mCubd} \in Dr_d$ 5. Update $D_{pdt} \leftarrow D_{pdt} + 1$ 6. Update $D_{dtmp}^{c5.0} \leftarrow C_{RrmRg}^{P5.0}(Rr_{1mCubrd}, d)$ 7. End For 8. Update $D_d^{c5.0} \leftarrow D_d^{c5.0} + D_{dtmp}^{c5.0}$ 9. End For 10. Estimate final result $fc_d^{c5.0} \leftarrow C_{class}^{c5.0}(D_d^{c5.0}, Sr_d)$

Figure 4: BI oriented data analysis and mining using C5.0 decision tree classification algorithm

For accomplishing the Sr_d data analysis the ultimate classification function $fc_d^{c5.0}$ can be further integrated with data warehouse or OLAP application that would result into optimized performance for accurate and precise decision making process in Business intelligence (BI) tools.

V. OLAP AND DATA MINING INTEGRATION

The process of effective data mining and information exploration from datasets or warehouse relate to the phenomenon of exploring certain genuine, unique, till unknown and potentially significant presentations of information from collected unprocessed datasets available in huge cubes or databases. The resemblance of the process called “Data mining” advocates the sifting in the course of huge raw data so as to achieve certain valuable information. In fact, it presents a multi- phase, recursive inductive phenomenon that encompasses processes like analysis of issues, extraction and exploitation of huge data, data preparation and its cleaning, reduction of unwanted data, rule generation, results analysis and prediction or decision making. Data warehouse or Online Analytical Processing (OLAP) represents an approach of presenting expected information from certain data bases or multidimensional cubes formed. OLAP scheme can be employed for visualizing multidimensional data cubes and identifying the expected patterns, data orientation a relationship can be built. The process of visualization may encompass slicing and dicing among data where the authority or analytic may chose the significant data for accomplishing certain specific decisive information. It may re-orientation of data view, increment or decrease in abstraction level and filtering of data available. Hence, the multidimensional privacy preserved cubes, would facilitate data warehouse or OLAP to exhibit data analysis and decision making. In our work, once the privacy preserved and enhanced decision tree mining framework has been developed, the concept hierarchies has been employed to perform generalization of individual node in decision tree that can be further accessed by data warehouse applications or OLAP functions. Thus, such unique system implementation makes the overall business intelligence (BI) process more robust and efficient for service oriented applications for varied industries or organizations.

VI. RESULTS AND ANALYSIS

In this paper, the proposed system model was developed with robust C5.0 decision tree classification algorithm so as to achieve optimum performance results in mining process that may further enhance the overall efficiency of certain Business intelligence application. The proposed system model has been developed on Visual Studio platform and C#, .net programming language has been employed for it. Since, data mining is one of the most predominant factors that optimize business intelligence (BI) optimization. For optimizing classification accuracy and rule generation, C5.0 has been developed which was further enhanced with its privacy preservation features. In order to justify the robustness of proposed C5.0 algorithm based framework, we have developed a parallel reference model using C4.5 decision tree algorithm. In comparative study, on the basis of results analysis for rule set generation, testing accuracy, training accuracy and classification aspects, the proposed C5.0 algorithm based model has over performed its descendant C4.5 classification based mining model or system.

The implementation of proposed system model and its comparative study has been done in terms of privacy preserved rule set size and rule generation effectiveness, accuracy analysis for classification rule based training process and similarly, accuracy analysis for testing and area covered in computation. The retrieved data illustrates that the proposed C5.0 based decision tree classification performs better with data warehouse/OLAP frameworks or business intelligence tools is given in **Figure 5 & 6** as below:

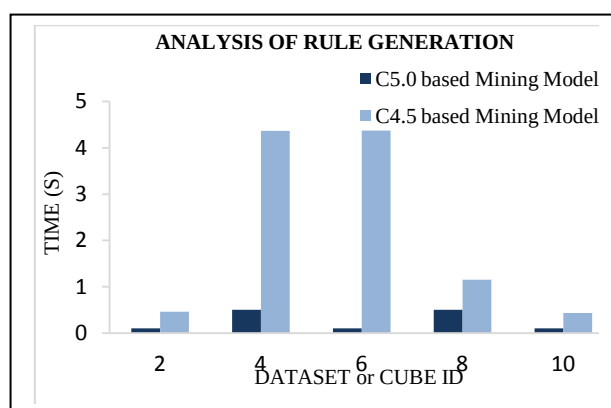


Figure 5: Analysis for rule generation

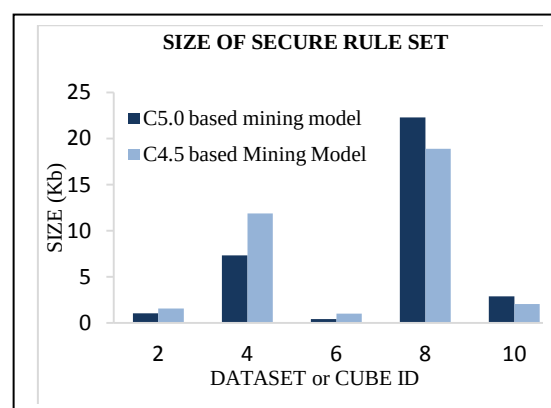


Figure 6: Analysis for rule set generation

As depicted in Figure7 to 10 as below, the factor area under curve for the framework analysis might be considered as the accuracy approximation toward positive instances. This characteristic exhibits that the probability of classifier may accomplish higher positive occurrence in comparison with the randomly selected negative instances. This also depicts that proposed C5.0 encompasses minimum computational overheads and cost that might strengthen the OLAP or data warehouse frameworks for business intelligence applications or tools.

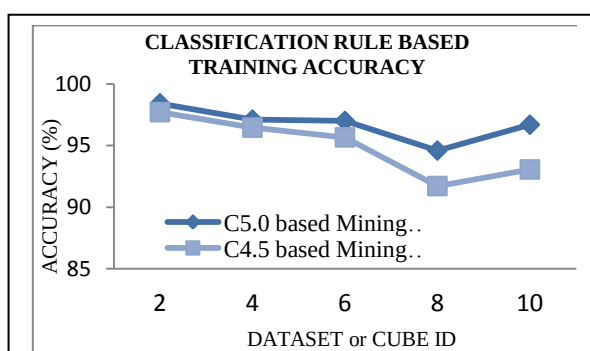


Figure 7: Classification training accuracy analysis

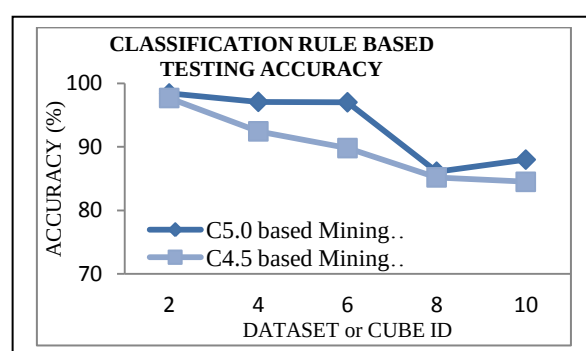


Figure 8: Classification testing accuracy

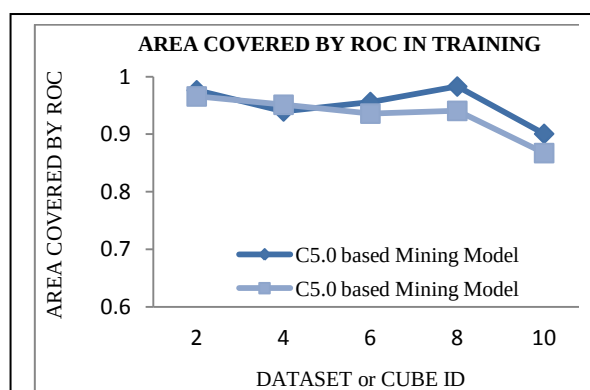


Figure 9: Comparative analysis for area covered by covered by C4.5 and C5.0

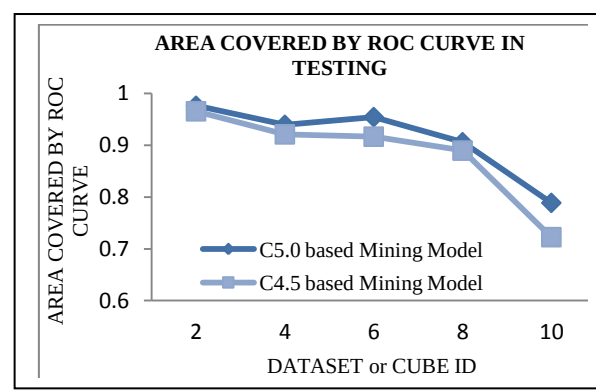


Figure 10: Comparative analysis for area ROC for data set testing

This enhanced performance can enrich decision support system (DSS) with higher accuracy and minimum complexity even with the huge data set's processing.

VII. CONCLUSION

Because of the huge data volume in certain organization or institutions, business intelligence (BI) becomes very significant process for decision making and optimistic strategic planning. Similarly data warehouse and online analytical processing (OLAP) also have vital significances for business intelligence (BI) applications and decision support systems (DSS). Data mining is one of the key components of data warehouse and BI models. The optimization of decision tree classification in data mining model can optimize the overall performance of business intelligence (BI) model. In this paper, a highly robust and privacy preserved, data centric mining module has been developed for BI applications. Considering the need of efficient classification and rule generation approach in data mining, in this work C5.0 decision tree classification algorithm has been employed

to perform mining which ultimately enhances the performance of integrated BI framework. The developed mining model has exhibited better results in terms of rule set generation, data classification, learning/training accuracy with enhanced optimized area covered by receiver operating characteristics. The proposed C5.0 algorithm has outperformed C4.5 based mining model. The implementation of proposed mining model with OLAP/data warehouse applications and BI frameworks would result into highly efficient, privacy preserved and most precise decision making process for BI applications.

REFERENCES

- [1] Z. Michalewicz; M. Schmidt; M. Michalewicz, C. Chiriac; "Adaptive Business Intelligence"; Springer, 2006
- [2] L. Nemuraite; L. Ceponiene; G. Vedrickas; "Representation of Business Rules in UML&OCL Models for Developing Information Systems" Lecture Notes in Business Information Processing in 2009, 182-196.
- [3] nmon, WH; Building the data warehouse; John Wiley in 1992.
- [4] W. W. Eckerson; "Performance Dashboards: Measuring, Monitoring and Managing Your Business"; Wiley, 2010.
- [5] J. Ranjan; "Business justification with business intelligence"; I Vine, pp. 461-475, 2008.
- [6] U. Sekaran; "Research Methods for Business"; John Wiley & Sons in 2006.
- [7] O. Goldreich; "Secure multi-party computation"; Final draft, version 1. 4 in 2002.
- [8] C. Dwork; K. Kenthapadi; F. McSherry; I. Mironov; and M. Naor; "Our data, ourselves: Privacy via distributed noise generation"; Advances in Cryptology-EUROCRYPT; pages 486-503, 2006.
- [9] Xindong Wu; Vipin Kumar; J. Ross Quinlan; Joydeep Ghosh; Qiang Yang; Hiroshi Motoda; Geoffrey J. McLachlan; Angus Ng; Bing Liu and Philip S. Yu, et al. "Top 10 algorithms in data mining"; Springer; Knowledge and Information Systems Volume 14, Number 1 in 2008.
- [10] Tomasz Bujlow; Tahir Riaz; Jens Myrup Pedersen; "A method for classification of network traffic based on C5.0 Machine Learning Algorithm" in workshop on Computing, Networking and Communications in 2012 IEEE pp 237-241.
- [11] Po-Hsun. Sung; Jyh-Dong Lin; Shih-Huang Chen; Shun-Hsing Chen; and Jr-Hung Peng; "Utilization of Data Mining on Asset Management of Freeway Flexible Pavement"; Proceedings of IEEE IEEM; Pp 977 - 979 in 2010.
- [12] R. Agrawal and R. Srikant; "Privacy preserving data mining"; in Proc. ACM SIGMOD Int'l Conf. on Management of Data in 2000.
- [13] Guarda, T.; Santos, M.F.; Augusto, M.F.; Silva, C.; Pinto, F.; "Process Mining: A framework proposal for Pervasive Business Intelligence," Information Systems and Technologies (CISTI), 8th Iberian Conference on 19-22 June 2013, pp.1-4.
- [14] Yang Hang; Fong, S., "Real-time business intelligence system architecture with stream mining," Digital Information Management (ICDIM), 2010 Fifth International Conference on 5-8 July 2010, pp.29-34.
- [15] Yang Hang; Fong, S., "A Framework of Business Intelligence-Driven Data Mining for E-business," INC, IMS and IDC, Fifth International Joint Conference on 25-27 Aug. 2009, pp.1964-1970.
- [16] Ping-Tsai Chung; Chung, S.H., "On data integration and data mining for developing business intelligence," Systems, Applications and Technology Conference (LISAT) on 3-3 May 2013, pp.1-6.
- [17] Martin, A.; Lakshmi, T.M.; Prasanna Venkatesan, V., "An analysis on business intelligence models to improve business performance," Advances in Engineering, Science and Management, International Conference on 30-31 March 2012, pp.503-508.
- [18] Martin, A.; Miranda Lakshmi, T.; Venkatesan, V.Prasanna, "A business intelligence framework for business performance using data mining techniques," Emerging Trends in Science, Engineering and Technology (INCOSSET), 2012 International Conference on 13-14 Dec. 2012, pp.373-380.
- [19] Ma Jun, "Research of detecting e-business fraud based on data mining," Computer Science and Network Technology, 2011 International Conference on 24-26 Dec. 2011, vol. no. 4, pp.2201-2204.
- [20] Ali, K.; Warraich, M.A., "A framework to implement data cleaning in enterprise data warehouse for robust data quality," Information and Emerging Technologies (ICIET), 2010 International Conference on 14-16 June 2010, pp.1-6.
- [21] Usman, M.; Pears, R., "A methodology for integrating and exploiting data mining techniques in the design of data warehouses," Advanced Information Management and Service (IMS), 2010 6th International Conference on Nov. 30 -Dec. 2 2010, pp.361-367.
- [22] Zare-Mirakabad, M.-R.; Kaveh-Yazdy, F.; Tahmasebi, M., "Privacy preservation by k-anonymizing Ngrams of time series," Information Security and Cryptology (ISCISC), 2013 10th International ISC Conference on 29-30 Aug. 2013, pp.1-6.
- [23] Jiandang Wu; Jiyi Wang; Jianmin Han; Hao Peng; Jianfeng Lu, "An anonymized method for classification with weighted attributes," Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on 5-8 Aug. 2013, pp.1-5.
- [24] Weiwei Fang; Bingru Yang; DingLi Song; Zhigang Tang, "A New Scheme on Privacy-Preserving Distributed Decision-Tree Mining," Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on 7-8 March 2009, vol. no. 2, pp.517-520.
- [25] Mumtaz, S.; Rauf, A.; Khusro, S., "A distortion based technique for preserving privacy in OLAP data cube," Computer Networks and Information Technology (ICCNIT), 2011 International Conference on 11-13 July 2011, pp.185-189.
- [26] KumaraSwamy, S; Manjula S H; Venugopal, K R; Iyengar S S; Patnaik, L M, "Association rule sharing model for privacy preservation and collaborative data mining efficiency," Engineering and Computational Sciences (RAECS) on 6-8 March 2014; pp.1-6.

ACKNOWLEDGMENT

One of the author's Mr. Rajasekharaiah K.M. thanks Ms. Chhaya Dule, Asst.Prof. Jyothy Institute of Technology, Bangalore for her valuable suggestions.

AUTHOR:

Presently Mr. **Rajasekharaiah K.M.** is working as Professor & HOD Department of Computer Science & Engineering, Jnana Vikas Institute of Technology, Bangalore. He has done M.Tech. in Computer Science & Engg. M.Sc. Information Technology, M.Phil. in Computer Science, and PGDIT from reputed Universities, India. He is having 30+ years of total experience including 16 years of Industrial experiences. He is a Life fellow Member of Indian Society for Technical Education (ISTE), New Delhi. He is presently pursuing the doctoral degree in the Branch of Computer Science & Engineering, in the domain area of Data Mining & Warehousing. He has research publications in reputed national and international journals. His other area of interests are DBMS, Software Engg., Software Architecture, Computer Networks, Programming Languages, Data Structures and Mobile Computing. He is also a resource scholar for other Engineering Colleges/University.

