Placement Prediction Analysis in University Using Improved Decision Tree Based Algorithm

Dammalapati Rama Krishna[*1], Bode Prasad[#2], Teki Satyanarayana Murthy[#3]

[1#] *M.Tech , Department of Information Technology, Vignan's Institute of Information Technology,Visakhapatnam, Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India*

[2#] *Associate Professor, Department of Information Technology, Vignan's Institute of Information Technology, Visakhapatnam, Andhra Pradesh, India.*

3# Assistant Professor, Department of Computer Science and Engineering, Vignan University Vadlamudi, Guntur, Andhra Pradesh, India

***Abstract*—Data mining is an emerging research area applied where researchers applies for science and business application. In this paper data mining techniques are apply to placement prediction where recruitment taken place. Recruitment is one of the most important functions for any organization as they seek talented and qualified professionals to fill up their positions. Majority of the companies have been focusing on campus recruitment to fill up their positions. This method is the best way to get the right resources at the right time in the corporate world the young budding engineers. The focus of this paper is to identify whether the student will get placement or not. While the industry get the best talent from different institutes/universities, students too get chance to start their career with some of the best but it is very difficult in getting the placements. The result of this paper will assist will improve the performance of students in terms of placement. By applying the Improved Decision Tree Based classification algorithms on this data, we have predicted that students will placed in Recruitment Drives.**

***Keywords*—Classification, Decision Tree, Data Mining, Educational Research, Placement, Predicting Performance, Decision tree**

I. INTRODUCTION

Data mining is widely used by researchers for science and enterprise purposes. Sample data sets collected from individuals are necessary for decision making. In order to keep the utility of samples privacy preserving process is developed to sanitize when their private information is given to third party for processing or computing.[1],[2] During the storing process, samples can be stolen or leaked any time or while residing in storage This paper concentrates on preventing such attacks on third parties for the whole life time samples. Recruitment is one of the most important functions for any organization as they seek talented and qualified professionals to fill up their positions. Majority of the companies have been focusing on campus recruitment to fill up their positions. This method is the best way to get the right resources at the right time with minimal companies in the corporate world at the beginning of their career.. This can very well be achieved using the concepts of data mining. For this purpose, we have analyzed the data of students [3] engineering. This data was obtained from the information provided by the admitted students to the institute. It includes their S.No, city, communication skills, Technical skills, grade, attitude, economical background, written test,. We then applied the Improved Decision Tree Based Algorithm after pruning the dataset to predict the results of these students in their as precisely as possible. This paper covers the application of this new privacy preserving approach with the [4] Decision tree learning algorithm both continuous and discrete valued attributes only.

II. RELATED WORK

Data mining is the process of discovering interesting knowledge, such as associations, [5],[6]patterns, changes, significant structures and anomalies, from large amounts of data stored in databases or data warehouses or other information repositories It has been widely used in recent years due to the availability of huge amounts of data in electronic form, and there is a need for turning such data into useful information and knowledge for large applications. These applications are found in fields [7],[8],[9]such as Artificial Intelligence, Machine Learning, Market Analysis, Statistics and Database Systems, Business Management and Decision Support Classification is a data mining technique that maps data into predefined groups or classes. It is a supervised learning method which requires labeled training data to generate rules for classifying test data into predetermined [10],[11],[12]groups or classes It is a two-phase process. The first phase is the learning phase, where the training data is analyzed and classification rules are generated. The next phase is the classification, where test data is classified into classes according to the generated rules. Since classification algorithms require that classes be defined based on data attribute values, we had created an attribute "class" for every student, which can have a value of either "Pass" or "Fail". In clustering, classes are unknown apriori and are discovered from the data. [13],[14]Since our goal is to predict

students' performance into either of the predefined classes - "Pass" and "Fail", clustering is not a suitable choice and so we have used classification algorithms instead of clustering algorithms. Issues Regarding Classification: Missing data values cause problems during both the [15],[16] training phase and to the classification process itself. For example, the reason for non-availability of data may be due to Equipment malfunction, Deletion due to inconsistency with other recorded data ,Non-entry of data due to misunderstanding, certain data considered unimportant at the time of entry, No registration of data or its change, Data miners can ignore the missing data, Data miners can replace [17],[18],[19]all missing values with a single global constant, Data miners can replace a missing value with its feature mean for the given class, Data miners and domain experts, together, can manually examine samples with missing values and enter a reasonable, probable or expected value .In our case, the chances of getting missing values in the [20]training data are very less. The training data is to be retrieved from the admission records of a particular institute and the attributes considered for the input of classification process are mandatory for each student. The tuple which is found to[21],[22],have missing value for any attribute will be ignored from training set as the missing values cannot be predicted or set to some default value. Considering low chances of the [23], occurrence of missing data, ignoring missing data will not affect the accuracy adversely. The measuring accuracy determining that which data mining technique is best depends on the interpretation of the problem by users. Usually, the performance of algorithms is examined by evaluating the accuracy of the result. Classification accuracy is calculated by determining the percentage of tuples placed in the correct class. At the same time there may be a cost associated with an incorrect assignment to the wrong class which can be ignored.

III. ID 3 Decision Tree Algorithm

The ID3 algorithm is an extension of the ID3 algorithm, and proposed by Quinlan. After years of improvement, ID3 algorithm is one of the best algorithms in handling numeric attributes.It finds the best splitting attribute and the best splitting point of the numeric continuous attributes. Here the attribute with maximum gain ratio is selected as the splitting attribute.

Attribute selection measure

The information gain is measure is used to select the test attribute at each node in the tree. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split. After calculating information gain of each attribute we chose test attribute for the current node. Among the attributes which has the highest information gain (or greatest entropy reduction).Let S be a set consisting of s data samples. Suppose the class label attribute has m distinct classes, C_i(for i = 1…m). Let $s_i$be the number of samples of S in class C_i.the expected information needed to classify a given sample is given by

$$I(s_1\, s_2\, , \dots \dots , s_m) = -\sum_{i=1}^{m} p_i \log_2 (p_i)$$

Where p_i is the probability that an arbitrary sample belongs to class C_I and is estimated by s/s. The log function to the base 2 is used as the information is encoded in bits.Let attribute A have v distinct values,$\{a_1,a_2 ,\dots..a_v\}$. attribute A can be used to partition S into v subsets, $\{S_1,S_2,\dots\dots\dots S_V\}$,where S_j contains those samples in S that have value $a_j$of A. if A were selected as the test attribute (the best attribute for splitting), then those subsets would correspond to the branches grown from the node containing the set S. let s_{ij}be the number of samples of class C_I in a subset S_j . the entropy, or expected information based on the partitioning into subsets by A , is given by

$$E(A)=\sum^{V}_{j=1}(S1j+\dots+Smj)/S*I(S1j+\dots+Smj)$$

The term (S1j+…+Smj)/S acts as the weight of j^{th} subset and is the number of the jth subset and is the number of samples in the subset (halving value $a_j$of A)divided by total number of samples in S. the smaller the entropy value, the greater the purity of the subset partitions. For a given subset S_j,

$$I(S1j,S2j,\dots\dots.,Smj)=-\ \sum^{m}_{i=1}P_{ij}\log P_{ij}$$

*I*Where $P_{ij=}S_{ij}/|S_j|$ and is the probability that a sample in S_j belongs to class C_j. The encoding information that would be gained by branching on A is

$$Gain(A)=I(S1,S1,\dots.Sm)-E(A)$$

In other words, Gain (A) is the expected reduction in entropy caused by knowing the value of attribute A.

$$Gain\ ratio(A)=gain(A)/split\_info(A)$$

$$Split\ info(A)=\sum V_{i=1}\ (S1j+\dots+Smj)/S$$

For each attribute calculate the gain ratio by using the computational algorithm then chosen the best attribute with the highest gain ratio is chosen as the testing attribute for the given dataset S .

IV. Decision Tree Generation

In this algorithm selects a test attribute according to the information content of the training set T_S. Majority value retrieves the most frequent value of the decision attribute of T_S. Construct the tree with root element best value. Generate a sub tree with T_S and attribute best. Connect tree and sub tree to generate the decision tree. Decision tree learning is the process of inducing a decision tree from a training set T. the decision tree G is built by the top-down approach recursively, starting from the root node.

Algorithm: Generate – tree (T_S, attribs, default)
Input: T_S, the set of training data sets attribs, set of attributes, Default, default value of the goal predicate
Output: Tree, a decision tree

1. If T_S is empty then return default
2. Default ←majority – value (T_S)
3. If $H_{ai}(T_S)$=0 **then return** default
4. **Else if** attribs is empty **then return** default
5. Else
6. Best ← choose –attribute (attribs, T_S)
7. Tree ← a new decision tree with root attribute best
8. For each value v_i of best **do**
9. T_{si}← {datasets in , T_S as best =k_i}
10. Subtree ←generate –tree (T_S, attribs –best , default)
11. Connect tree and subtree with a branch labelled k_i
12. Subtree ←generate –tree (T_S, attribs –best , default)
13. Connect tree and subtree with a branch labelled k_i
14. Return tree

Generate –tree process by applying the ID3approach with the original samples T_S

Entropy (s) = $\in$-P(I) $\log_2$ P(I)

Entropy(s)=(63/86)$\log_2$(63/86)–(23/86) $\log_2$ (23/86)= 0.14001

Economical Background: Among 86 students, 69 RICH students out of that 51 are not placed and 18 students are placed and the remaining are 17 POOR students, 12 are not placed and6 students are placed. Among 86 students, 72 GOOD students out of that 54 are not placed and 18 students are placed and the remaining are 14 BAD students, 9 are not placed and 5 students are placed.

Gain(S,EB)=Entropy(S) –(63/86) * Entropy (SRICH) (23/86) * Entropy (SPOOR)

Entropy (E Rich) = (51/69) $\log_2$ (51/69) – 18/69) $\log_2$ (18/69) = 0.14397

Entropy (E Poor) = (12/17) $\log_2$ (12/17) – (6/17) $\log_2$ (6/17) = 0.1124956

Gain (S, EB)=0.14- (63/86)*0.143-(23/86) *0.112 = 0.1049019

Attitude: Among 86 students, 72 Good students out of that 54 are not placed and 18 students are placed and the remaining are 14 Bad students, 09 are not placed and 05 students are placed.

Entropy (E, Good)= (54/72) $\log_2$ (54/72) – (18/72) $\log_2$ (18/72) = 0.150514

Entropy (E, Bad)=(9/14) $\log_2$ (9/14) – (5/14) $\log_2$ (5/14) = 0.08600

Gain (S, ATTITUDE) = 0.14 – (63/86) * 0.15 – (23/86) * 0.08= 0.006752

Technical Skills: Among 86 students, 53 Good students out of that 30 are not placed and 23 students are placed and the remaining are 33 Bad students, 33 are not placed and 0 students are placed.

Entropy (S Good) = (30/53) $\log_2$ (30/53) – (23/53) $\log_2$ (23/53) = 0.039758

Entropy (S Bad) = (33/33) $\log_2$ (33/33) – (0/33) $\log_2$ (0/33) = 0.301029

Gain (S Tech) = 0.14001 – (63/86) * 0.039758 – (23/86) * 0.3010 = 0.0303

Communication Skills: Among 86 students, 56 Good students out of that 33 are not placed and 23 students are placed and the remaining are 33 Bad students, 30 are not placed and 0 students are placed.

Entropy (E Good) = (30/56) $\log_2$ (33/56) – (23/56) $\log_2$ (23/56) = 0.05375

Entropy (E Bad) = (30/30) $\log_2$ (30/30) – (0/30) $\log_2$ (0/30) = 0.301029

Gain(S,CS)=0.1401(63/86)*0.0535(23/86)*0,3010=0 .0201

Written Test: Among 86 students, 61 qualified students out of that 38 are not placed and 23 students are placed and the remaining are 25 Bad students, 25 are not placed and 0 students are placed.

Entropy (E Qualified)= (38/61) $\log_2$ (38/61) – (23/61) $\log_2$ (23/61) = 0.07402

Entropy (E not Qualified)= (25/25) $\log_2$ (25/25) – (0/25) $\log_2$ (0/25) = 0.3010

Gain(S, Written Test)= 0.14001 – (63/86) * 0.07402 – (23/86) * 0.301029 = 0.005278

City: Among 86 students, 39 Guntur Qualified students out of that 29 are not placed and 10 students are placed and the remaining are 12 Hyderabad students, 9 are not placed and 3 students are Vijayawada students, 11 are not placed and 6 students are placed and the remaining are 9 Kurnool students, 6 are not placed and 3 students

Entropy (E Guntur) = (29/39) $\log_2$ (29/39) – (10/39) $\log_2$ (10/39) = 0.14665

Entropy (E Hyderabad)= (9/12) $\log_2$ (9/12) – (3/12) $\log_2$ (3/12) = 0.15051

Entropy (E Vizag) = (8/9)$\log_2$(8/9) (1/9) $\log_2$ (1/9)= 0.23

Entropy (E Vijayawada) = (11/17) $\log_2$ (11/17) – (6/17) $\log_2$ (6/17) = 0.08853

Entropy (E Kurnool) = (6/9)$\log_2$(6/9)–(3/9)$\log_2$ (3/9) =0.10

Gain (S.City) = 0.14001 – (63/86) * 0.14665 – (23/86) * 0.15051* 0.23413*0.08853 * 0.10034 = **0.032496**

Grade: Among 86 students, 32 A Grade students out of that 18 are not placed and 14 students are placed and the remaining are 33 B Grade students, 27 are not placed and 6 students are placed. and the remaining are 21 C Grade students, 18 are not placed and 3 students are placed.

Entropy (E A Grade)= (18/32) $\log_2$(18/32) – (14/32) $\log_2$(14/32) = 0.03

Entropy (E B Grade) = (27/33) $\log_2$(27/33) – (6/33) $\log_2$(6/33)= 0.1915
Entropy (E C Grade)= (18/21) $\log_2$(18/21) – (3/21) $\log_2$(3/21) = 0.2150
Gain(S, Grade) = 0.14001 – (63/86) * 0.03762 – (23/86) * 0.19156 * 0.21532 = 0.101435

V. IMPLEMENTATION

We had divided the entire implementation into five stages. The first stage, information about students who have been eligible for placement data was collected. This included the details of the students whether placed or not placed .In the second phase, the data . The third stage involved applying the Improved C4.5 Decision Tree Based Learning algorithms. Let T be the Training Set where attributes like 1-SNO 2-CITY 3-COMMUNICATION SKILLS 4-TECHNICAL SKILLS 5-GRADE 6-ATTITUDE 7-ECONOMICAL BACKGROUND 8-WRITTEN TEST 9-PLACEMENT and the values like GU-Guntur,VJ-VijayawadaV-Vizag,H-Hyderabad,K-KURNOOL,GO-Good,B-Bad,Q-Qualified,NQ-NotQualified, N-No,Y-Yes,R-Rich,PO-Poor.

The below table illustrates the output simulation of university students

1	2	3	4	5	6	7	8	9
1	GU	GO	GO	A	GO	R	Q	Y
2	H	GO	GO	A	GO	R	Q	Y
3	GU	GO	B	B	GO	R	Q	N
4	GU	B	GO	A	GO	R	NQ	N
5	H	GO	B	B	GO	R	Q	N
6	V	B	GO	C	B	R	NQ	N
7	V	GO	B	A	GO	PO	Q	N
8	VJ	GO	GO	A	GO	R	Q	Y
9	GU	GO	GO	A	GO	R	Q	Y
10	V	GO	B	C	GO	R	Q	N
11	H	GO	B	B	GO	R	Q	N
12	H	GO	GO	A	GO	PO	Q	Y
13	H	B	GO	A	B	R	NQ	N
14	H	GO	GO	A	GO	R	Q	Y
15	H	GO	B	B	GO	R	Q	N
16	H	B	GO	A	GO	R	NQ	N
17	H	GO	B	B	GO	R	Q	N
18	H	GO	B	A	GO	R	Q	N
19	H	B	GO	A	GO	PO	NQ	N
20	H	GO	B	B	GO	R	Q	N
21	VJ	GO	GO	A	GO	R	Q	Y
22	VJ	B	GO	C	GO	PO	NQ	N
23	VJ	GO	B	B	GO	R	Q	N
24	VJ	B	GO	A	GO	R	NQ	N
25	VJ	GO	GO	B	GO	PO	Q	Y
26	VJ	GO	B	A	GO	R	Q	N
27	VJ	GO	GO	B	GO	R	Q	Y
28	VJ	B	GO	A	GO	R	NQ	N
29	VJ	GO	B	B	GO	R	Q	N
30	VJ	B	GO	C	GO	R	NQ	N
31	GU	GO	GO	A	GO	PO	Q	Y
32	GU	GO	B	B	GO	R	Q	N
33	GU	GO	B	C	B	R	Q	N
34	GU	B	GO	A	GO	PO	NQ	N
35	GU	GO	GO	A	GO	R	Q	Y
36	GU	GO	GO	B	GO	R	Q	Y
37	GU	GO	B	A	B	R	Q	N
38	GU	B	GO	B	GO	R	Q	N
39	GU	GO	B	C	GO	R	NQ	N
40	GU	GO	B	A	GO	PO	NQ	N
41	GU	B	GO	B	GO	R	Q	N
42	GU	GO	B	B	GO	R	NQ	N
43	GU	B	GO	C	GO	PO	Q	N
44	GU	GO	B	B	GO	R	NQ	N
45	GU	B	GO	C	GO	R	Q	N
46	GU	GO	GO	A	GO	PO	Q	Y
47	GU	B	GO	B	GO	R	Q	N
48	GU	GO	B	C	GO	R	NQ	N
49	GU	B	GO	B	GO	R	Q	N
50	GU	B	GO	C	GO	R	Q	N
51	GU	B	GO	B	B	R	Q	N
52	GU	GO	B	B	GO	R	NQ	N
53	GU	B	GO	B	GO	R	Q	N
54	GU	GO	GO	A	GO	R	Q	Y
55	GU	GO	B	C	GO	R	NQ	N
56	GU	B	GO	A	GO	R	Q	N
57	GU	B	GO	C	GO	R	Q	N
58	GU	GO	B	C	B	R	NQ	N
59	GU	GO	GO	B	GO	R	Q	Y
60	GU	GO	GO	A	GO	R	Q	Y
61	V	GO	B	A	GO	R	NQ	N
62	V	B	GO	C	GO	R	Q	N
63	V	GO	B	B	GO	R	Q	N
64	V	GO	GO	A	GO	R	Q	Y
65	V	GO	B	C	GO	R	NQ	N
66	V	B	GO	B	GO	R	Q	N
67	K	GO	GO	A	GO	R	Q	Y
68	K	B	GO	B	GO	R	Q	N
69	K	GO	B	B	GO	R	NQ	N
70	K	B	GO	B	GO	R	Q	N
71	K	GO	B	A	GO	R	NQ	N
72	K	B	GO	C	GO	R	Q	N
73	K	GO	B	B	GO	R	Q	N
74	K	GO	GO	A	GO	R	Q	Y
75	K	GO	GO	C	GO	R	Q	Y
76	GU	GO	GO	B	GO	R	Q	Y
77	GU	B	GO	B	B	R	Q	N
78	VJ	GO	B	A	GO	R	NQ	N
79	VJ	GO	B	C	B	R	Q	N
80	VJ	GO	GO	A	GO	R	Q	Y
81	VJ	GO	GO	A	GO	R	Q	Y

82	VJ	B	GO	C	GO	R	NQ	N
83	VJ	GO	B	A	GO	PO	Q	N
84	GU	B	GO	C	GO	R	NQ	N
85	GU	GO	B	B	GO	R	Q	N
86	GU	B	GO	C	B	PO	Q	N

VI. Modified Decision Tree Generation - Result

As of the Placement Prediction original data sets, T can be determined by the retrievable information from an university– the contents of training set T and modified decision tree generated from by the following algorithm. Applying the randomized parameter where @ is squared i.e 0<@<1.The @ randomized parameter effect the original Gain values of the attributes

Algorithm: Generate – Decisio tree (size T', attribs, default)

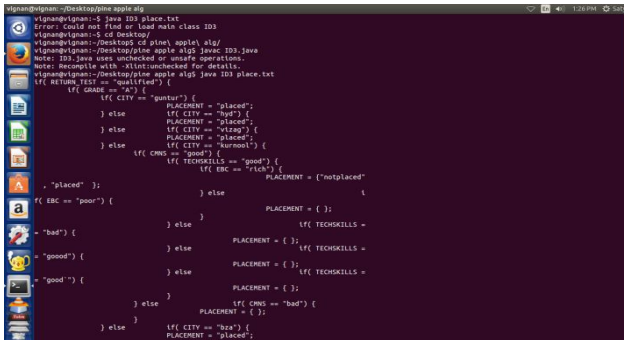
Input: size, T, the set of training data sets
T^P, the set of perturbing data sets, attribs, set of attributes

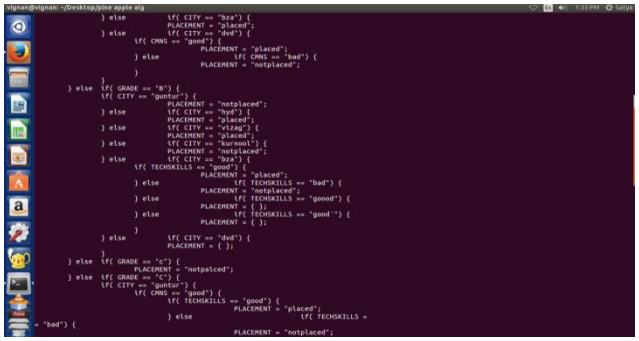
Default, default value for the goal predicate

Output: tree,a decision tree

1. If (T is empty **then return** default)
2. best ← choose Attribute (attribs, size, (T))
3. tree ← a new decision tree with root attribute best
4. size ← size/number of possible values k_i in best
5. Caluclate the gain by applying the randomized parameter
6. Subtree ← generate –tree (size, attribs –best, default)
7. Connect tree and sub tree with a branch labelled k_i
8. Return tree

SCREENSHOTS OF OBTAINED RESULT





VII. Conclusion

We proposed an approach gives the better results. This Improved decision tree learning algorithm handles the missing data, Continuous Data. Most widely used for classifying the large amounts of data easily, that helps in Prediction of a Student whether placed or not. This approach cannot handles have dynamic growing or shrinking facility based on the university norms.

VIII. Acknowledgment

We express sincere gratitude to our project guide, **Mr. Bode Prasad**, Associate Professor for his guidance and support. I consider it as a privilege to thank **Mr. Teki Satyanarayana Murthy**, Assistant Professor for helping me all through the project. I would like to thank our ever-inspiring Principal of Vignan's Institute of Information Technology **Dr. K. Alice Mary** for her encouragement to me during the work.

References

[1] Han, J. and Kamber, M., (2006) *Data Mining: Concepts and Techniques*, Elsevier.

[2] Dunham, M.H., (2003) *Data Mining: Introductory and Advanced Topics*, Pearson Education Inc.

[3] Kantardzic, M., (2011) *Data Mining: Concepts, Models, Methods and Algorithms*, Wiley-IEEE Press.

[4] Ming, H., Wenying, N. and Xu, L., (2009) "An improved decision tree classification algorithm based on ID3 and the application in score analysis", Chinese Control and Decision Conference (CCDC), pp1876-1879.

[5] Xiaoliang, Z., Jian, W., Hongcan Y., and Shangzhuo, W., (2009) "Research and Application of the improved Algorithm C4.5 on Decision Tree", International Conference on Test and Measurement (ICTM), Vol. 2, pp184-187.

[6] CodeIgnitor User Guide Version 2.14, http://ellislab.com/codeigniter/user-guide/toc.html

[7] RapidMiner, http://rapid-i.com/content/view/181/190/

[8] MySQL – The world's most popular open source database, http://www.mysql.com/

[9] PuiK.Fong and Jens H. Weber-Jahnke,SeniorMember,IEEE Computer Society,"Privacy Preserving Decision Tree Learning Using Unrealized Data Sets".

[10] S. Ajmani,R.morris, and B. Liskov,"A trusted third –party computation service," Technical Report MIT-LCS-TR-847.MIT,2001

[11] .LWangandA.Jafari,"HidingSenstitivePredictiveAssociationRules ," proc. IEEE int'l Conf.Systems,M an and Cybernetics,pp. 164-169,2005

[12] R.Agrawal and R.Srikant,"privacy preserving datamining,"proc.ACMSIGMODconf.management of data (SIGMOD '00),pp.439-450, may 2000.

[13]] Q.Ma and P Deng,"Secure Multi-Party protocols for Privacy Preserving Data mining ,"proc.ThirdInt'lConf.WirelessAlgorithm,Systems, and A pplications(WASA '08),pp.526-537,2008.

[14] J.Githanjali,J.Indumathi,N.C.Iyengar, and N.Sriman,"A Pristine clean Cabalistic ForutityStrategizeBased Approach for Incremental Data Stream Privacy Preserving Data Mining,"proc. IEEE Second Int'l Advance Computing Conf.(IACC),pp.410-415,2010

[15] N.Lomas,"Data on 84,000 United Kingdom Prisoners is Lost," Retrieved sept.12, 2008, http://news.cnet.com/8301-1009_3-10024550-83.html, aug. 2008.

[16]] BBC News Brown Apologises for Records Loss. Retrieved sept.12, 2008, http://news.bbc.co.uk/2/hi/uk_news/politics/7104945.stm,nov. 2007

[17] D.Kaplan,Hackers Steal 22,000 social Security Numbers From Univ. Of Missouri Database, Retrivedsept.2008, http://www.scmagazineus.com/Hackers-steal-22000-Social- Security -numbers-from-univ -of-Missouri-database/article/34964/,May 2007.

[18] D.Goodin,"Hackers Infiltrate TD Ameritrade CLIENT Database,"Retrived Sept2008, http://www.channelregister.co.uk/2007/09/15/ameritrade_database_burgled/, sept.2007.

[19] L.Liu, M kantarcioglu, and B. Thuraisingham, "Privacy Preserving Decision Tree Mining from Perturbed data ," proc. 42nd Hawaii Int'l conf System Sciences (HICSS '09),2009

[20] Y.Zhu,LHuang,WYang,DLi,Y.Luo,andF.Dong, "Three new approaches to privacy-preserving Add to multiply protocol and its application",Proc.second int'l workshop knowledge discovery and dataming,(WKDD'09),pp.554 558,2009

[21] C.Aggarwal and P.Yu, Privacy preserving data mining:,models and algorithms.Springer,2008.

[22] L. Shaneck and Y.Kim "Efficient cryptographic primitives for private datamining",Proc.43rd Hawaii int'l conf.systemsciences(HICSS),pp.1-9,2010.

[23] Varidyaand C. Clifton , "privacy preserving association rule mining in vertically partitioned data ", proc .Eighth ACM SIGKDD int'l Conf. Knowledge discovery and data mining(KDD '02),PP.23-26,July 2002