INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A., Google Scholar, Indian Citation Index (ICI), Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)], Index Copernicus Publishers Panel, Poland with IC Value of 5.09 (2012) & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than **5771 Cities** in **192 countries/territories** are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to the recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION**:

DATED: _____________

THE EDITOR

IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF ____.**

(e.g. Finance/Mkt./HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript titled '____' for likely publication in one of your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published anywhere in any language fully or partly, nor it is under review for publication elsewhere.

I affirm that all the co-authors of this manuscript have seen the submitted version of the manuscript and have agreed to inclusion of their names as co-authors.

Also, if my/our manuscript is accepted, I agree to comply with the formalities as given on the website of the journal. The Journal has discretion to publish our contribution in any of its journals.

NAME OF CORRESPONDING AUTHOR	:
Designation/Post*	:
Institution/College/University with full address & Pin Code	:
Residential address with Pin Code	:
Mobile Number (s) with country ISD code	:
Is WhatsApp or Viber active on your above noted Mobile Number (Yes/No)	:
Landline Number (s) with country ISD code	:
E-mail Address	:
Alternate E-mail Address	:
Nationality	:

* i.e. Alumnus (Male Alumni), Alumna (Female Alumni), Student, Research Scholar (M. Phil), Research Scholar (Ph. D.), JRF, Research Assistant, Assistant Lecturer, Lecturer, Senior Lecturer, Junior Assistant Professor, Assistant Professor, Senior Assistant Professor, Co-ordinator, Reader, Associate Professor, Professor, Head, Vice-Principal, Dy. Director, Principal, Director, Dean, President, Vice Chancellor, Industry Designation **etc**. ***The qualification of author is not acceptable for the purpose***.

NOTES:

a) The whole manuscript has to be in ***ONE MS WORD FILE*** only, which will start from the covering letter, inside the manuscript. ***pdf. version is liable to be rejected without any consideration***.

b) The sender is required to mention the following in the **SUBJECT COLUMN of the mail**:

New Manuscript for Review in the area of (e.g. Finance/Marketing/HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)

c) There is no need to give any text in the body of the mail, except the cases where the author wishes to give any **specific message** w.r.t. to the manuscript.

d) The total size of the file containing the manuscript is expected to be below **1000 KB**.

e) Only the **Abstract will not be considered for review** and the author is required to submit the **complete manuscript** in the first instance.

f) ***The journal gives acknowledgement w.r.t. the receipt of every email within twenty-four hours*** and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of the manuscript, within two days of its submission, the corresponding author is required to demand for the same by sending a separate mail to the journal.

g) The author (s) name or details should not appear anywhere on the body of the manuscript, except on the covering letter and the cover page of the manuscript, in the manner as mentioned in the guidelines.

2. **MANUSCRIPT TITLE:** The title of the paper should be typed in **bold letters**, **centered** and **fully capitalised**.

3. **AUTHOR NAME (S) & AFFILIATIONS:** Author (s) **name**, **designation**, **affiliation** (s), **address**, **mobile/landline number** (s), and **email/alternate email address** should be given underneath the title.

4. **ACKNOWLEDGMENTS:** Acknowledgements can be given to reviewers, guides, funding institutions, etc., if any.

5. **ABSTRACT:** Abstract should be in **fully Italic printing**, ranging between **150** to **300 words**. The abstract must be informative and elucidating the background, aims, methods, results & conclusion in a **SINGLE PARA**. ***Abbreviations must be mentioned in full***.

6. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of **five**. These should be arranged in alphabetic order separated by commas and full stop at the end. All words of the keywords, including the first one should be in small letters, except special words e.g. name of the Countries, abbreviations etc.

7. **JEL CODE:** Provide the appropriate Journal of Economic Literature Classification System code (s). JEL codes are available at www.aeaweb.org/econlit/jelCodes.php. However, mentioning of JEL Code is not mandatory.

8. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. *It should be free from any errors i.e.* ***grammatical, spelling*** *or* ***punctuation.*** *It must be thoroughly edited at your end*.

9. **HEADINGS:** All the headings must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

10. **SUB-HEADINGS:** All the sub-headings must be bold-faced, aligned left and fully capitalised.

11. **MAIN TEXT:**

THE MAIN TEXT SHOULD FOLLOW THE FOLLOWING SEQUENCE:

INTRODUCTION
REVIEW OF LITERATURE
NEED/IMPORTANCE OF THE STUDY
STATEMENT OF THE PROBLEM
OBJECTIVES
HYPOTHESIS (ES)
RESEARCH METHODOLOGY
RESULTS & DISCUSSION
FINDINGS
RECOMMENDATIONS/SUGGESTIONS
CONCLUSIONS
LIMITATIONS
SCOPE FOR FURTHER RESEARCH
REFERENCES
APPENDIX/ANNEXURE

The manuscript should preferably be in ***2000 to 5000 WORDS***, But the limits can vary depending on the nature of the manuscript.

12. **FIGURES & TABLES:** These should be simple, crystal **CLEAR**, **centered**, **separately numbered** & self-explained, and the **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. *It should be ensured that the tables/figures are referred to from the main text*.

13. **EQUATIONS/FORMULAE:** These should be consecutively numbered in parenthesis, left aligned with equation/formulae number placed at the right. The equation editor provided with standard versions of Microsoft Word may be utilised. If any other equation editor is utilised, author must confirm that these equations may be viewed and edited in versions of Microsoft Office that does not have the editor.

14. **ACRONYMS:** These should not be used in the abstract. The use of acronyms is elsewhere is acceptable. Acronyms should be defined on its first use in each section e.g. Reserve Bank of India (RBI). Acronyms should be redefined on first use in subsequent sections.

15. **REFERENCES:** The list of all references should be alphabetically arranged. ***The author (s) should mention only the actually utilised references in the preparation of manuscript*** and they may follow Harvard Style of Referencing. **Also check to ensure that everything that you are including in the reference section is duly cited in the paper**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc., in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italic printing. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parenthesis.
- ***Headers**, **footers**, **endnotes** and **footnotes** should **not be used** in the document.* However, **you can mention short notes to elucidate some specific point**, which may be placed in number orders before the references.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–23

UNPUBLISHED DISSERTATIONS

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

BAYESIAN NETWORKS STRUCTURE LEARNING USING CLASSIFICATION

HEENA TIMANI
ASST. PROFESSOR
SCHOOL OF COMPUTER STUDIES
AHMEDABAD UNIVERSITY
AHMEDABAD

Dr. MAYURI PANDYA
HEAD & PROFESSOR
DEPARTMENT OF STATISTICS
MAHARAJA KRISHNAKUMAR SINHJI BHAVNAGAR UNIVERSITY
VIDHYANAGAR

ABSTRACT

Data mining is becoming main stream technology used in business intelligence and applications. Data mining offers tools for discovery of relationship, patterns and knowledge from a massive database in order to guide decision about future activity. Probabilistic Graphical Models also known as Bayesian networks are popular and powerful tool in data mining. Bayesian networks provide a general and effective frame work for knowledge representation and reasoning under uncertainty. It is one of most effective theory models in expression of uncertainty knowledge because it has a strong ability for probabilistic reasoning and the characteristic of easy understanding to humans. A Bayesian network is a combination of a qualitative and a quantitative component. Structural information of domain can be represented in form of qualitative part and causality, relevance or (in) dependence relationships between variables. Using quantitative part, we can add uncertainty in to model and represents probability distributions that quantify these relationships. Once a complete Bayesian network has been built, it is an efficient tool for performing inferences. However, there still remains the previous problem of building such a network, that is, to provide the graphical structure and the numerical parameters necessary for characterizing it. As it may be difficult and time-consuming to build Bayesian networks using the method of eliciting opinions from domain experts, and there are various different domains that provides data. Learning Bayesian networks from this domain is challenging. Bayesian networks have received considerable attention from the machine learning community. There are many machine learning algorithms for automatically building Bayesian networks from data. In this paper algorithms based on Bayesian classification approach Bayes net, log score and structure learning are used for combinatorial optimization problem for commercial big data. Practical machine learning and data mining open source software weka 6.3.2 is used for knowledge discovery and data mining.

KEYWORDS

bayesian networks classifier, machine learning, scoring function, structure learning.

1. INTRODUCTION

Bayesian networks compactly represent a joint probability distribution by exploiting a set of independence assumptions between the set of random variables. A Bayesian network is a probabilistic graphical model where the random variables are represented as nodes and the conditional dependencies of the random variables are represented as edges. In Bayesian networks model selection and a structure that maximizes the posterior probability with respect to the data is very difficult. The space of all structures is super-exponential in the number of random variables and finding the optimal structure is known to be NP-hard. Many methods have been devised to overcome the obstacle of searching through such a large space. To cut down the space size greedy methods and decomposability of local scoring matrix like BDeu score. With decomposable metrics, the chosen modification can be evaluated separately from the rest of the network.

In traditional score-based search algorithms, the evaluation of the modification still requires looking at the entire dataset. If the size of data set increase the scanning database becomes very expensive.

Using a Bayesian network capturing the relationships between our uncertain beliefs in the propositions learning the truth value of one or more of the propositions, we can use Bayesian inference algorithms to find updated beliefs for each of the other propositions, and updated relationships between the propositions.

LEARNING BAYESIAN STRUCTURE

In technical term Bayesian network encodes a joint probability distribution over a set of random variables. A variable may be countable number or finite number of states or it may be continuous. Lower case letters are used to represent single variables and upper case letters are used to represent set of variables. We write x=m to denote the variable x is in state m. When we observe the state of every variable in set X, we call this set of observations a state of X. The joint space of a set of variables U is the set of all states of U. Using p(x/y) to denote the set of joint probability distribution over U is the probability distribution over the joint space of X, each one conditional on every state in the joint space of Y. A problem domain is a set of variables.

The structure of Bayesian network will depend on how the variables are ordered in the expansion of the equation 1. If the order is chosen carelessly, the resulting network structure may fall to reveal many conditional independencies in the domain. In practice, generally domain expert often can readily assert casual relationship among variables in a domain and use these assertions to construct a Bayesian network structure, without preordering the structure. Knowledge about uncertain domain is represented by graphical structure. each node in the graph represents a random variable and the edges between two nodes represent probabilistic dependencies among the corresponding random variables.

Conditional dependencies in the graph are estimated by known statistical and computational methods. Thus Bayesian networks combine ideology from graph theory, probability theory, computer science and statistics. Bayesian networks correspond to another Graphical Model structure known as a directed acyclic graph (DAG) that is popular in the Statistics, the Machine Learning. Bayesian Learning is both mathematically thorough and intuitively understandable. They enable an effective representation and computation of the joint probability distribution over a set of random variables (Pearl, 2000).

The structure of a directed acyclic graph is defined by two sets the set of nodes (vertices) and the set of directed edges. The nodes represent random variables and are drawn as circles labeled by the variables names. Direct dependence among the variables edges between nodes represented by arrows. The domain knowledge allows experts to draw an arc to a variable from each of its direct causes. Bayesian networks that specified the Joint Probability distribution using factored form, all possible inference queries by marginalization evaluated, i.e., summing out over 'irrelevant' variables. Two types of inference support are often considered: predictive support for node Xi, based on evidence nodes connected to Xi through its parent nodes (called also top-down reasoning) and diagnostic support for node, based on evidence nodes connected to Xi through its children nodes (called also bottom-up reasoning). In general, the full summation (or integration) over discrete (continuous) variables is called exact inference and known to be an NP-hard problem. Some efficient algorithms exist to solve the exact inference problem in restricted classes of networks. In many practical settings the Bayesian networks is unknown and one needs to learn it from the data. This problem is known as the Bayesian networks learning problem, which can be stated informally as follows.

Estimate the graph topology and the parameters of the joint probability distribution in the Bayesian networks. Node X_i has no parents then its local probability distribution is said to be unconditional. If the variable represented by a node is observed, then the node is said to be an evidence node, otherwise the node is said to be hidden node. In order to do Bayesian inference, prior probabilities and posterior probabilities are required. Each node is associated with it the conditional distribution for its variable. Each node has incoming edges from the nodes associated with the variables on which the node's conditional distribution is conditional Such a representation is a Bayesian network.

Several scoring functions for learning Bayesian networks have been proposed in the literature. It is common to classify scoring functions into two main categories: Bayesian and information-theoretic. In general, for efficiency purposes, these scores need to decompose over the network structure. The decomposability property allows for efficient learning algorithms based on local search methods. Moreover, when the learning algorithm searches in the space of equivalence classes of network structures, scoring functions must also be score equivalent, that is, equivalent networks must score the same.

2. CLASSIFICATION METHODOLOGY

Classification is an extensively researched topic in data mining and machine learning. The main hurdle to leveraging the existing classification methods is that these assume record data with a fixed number of attributes. In general, people dealing with sequence data use to convert the data into non-sequential data and then apply traditional classification algorithm. Commonly used classification algorithms are decision trees, k nearest neighbors, support vector machines, and bayes classifiers. Here, we briefly outline the basic classification algorithms. Though the methods to achieve classification vary based on the dataset and the model used, all the classifiers have something in common. The commonality is that they divide the given object space into disjunctive sections that are mapped to a given class. Bayesian classifiers are based on the assumption that the objects of a class can be modeled by a statistical process. Each data object has its origin the process of a given class with a certain probability and each process of a class generates objects with a certain probability called prior probability. To decide which class is to be predicted for a given object, it is necessary to determine the probability called the posterior probability of the object. It describes the probability that an object has its origin in that particular class. To determine the posterior probability, the rule of Bayes is used.

In Bayesian, classifier training is very fast. Also the model designed is simple and intuitive. Error is minimized in Bayesian, classifier subject to the assumptions of independence of attributes and data satisfying distribution model. Both these methods are used in machine learning and related algorithm know as Naive Bayes Classifier and Decision tree J48 are developed in weka 3.6.2 open source software. These two algorithms are used for classification of the data.

For classification and prediction decision trees are powerful and popular tools. Decision trees are more popular because in contrast to the neural network based approach they generate rules. These rules can easily be interpreted so that we can understand them. These rules can also be transformed into a database access language like SQL in order to retrieve quickly the records falling into a particular category.

For classification, the attribute values of a new object are tested beginning with the root. At each node, the data object can pass only one of the tests that are associated with the departing edges. The tree is traversed along the path of successful tests until a leaf is reached. Multiple algorithms have been proposed in the literature for constructing decision trees Generally, these algorithms split the training set recursively by selecting an attribute. The best splitting attribute is determined with the help of quality criteria. Examples of such quality criteria are information gain, gini index, Shannon information theory, and statistical significance tests. The advantages of decision trees are that they are very robust against attributes that are not correlated to the classes because those attributes will not be selected for a split. Another more important feature is the induction of rules. Each path from the root to a leaf provides a rule that can be easily interpreted by a human user. Thus, decision trees are often used to explain the characteristics of classes.

2.1 BAYESIAN SCORING FUNCTION

Compute the posterior probability distribution, starting from a prior probability distribution on the possible networks, conditioned to data T, that is, P (B|T). The best network is the one that maximizes the posterior probability. Since the term P (T) is the same for all possible networks, in practice, for comparative purposes, computing P(B, T) is sufficient. As it is easier to work in the logarithmic space, the scoring functions use the value log (P(B, T)) instead of P(B, T).

2.2 LOCAL SCORING METRICS

Local score metrics implemented Bayes, BDe, MDL, entropy, AIC. The minimum description length principle establishes an appropriate trade-off between complexity and precision in order to represent the network, we must store its probability values, and this requires a length which is proportional to the number of free parameters of the factorized joint probability distribution.

We have implemented various local search hill climbing algorithms such as K2, Hill Climb and Tree Augmented Naive Bayes. K2 adds arcs with a fixed ordering of variables. Hill climbing adding and deleting arcs with no fixed ordering of variables. Calculating the maximum weight spanning tree using Chow and Liu algorithm a tree can be formed.

3. DATA ANALYSIS AND RESULTS

Reserve bank data of year 1992 to 2012 is used for Bayesian classification Bayesian net structure learning and analysis of assets and liability. There are 45 attributes and 1859 instances. In the data one of the attribute is bank groups. There are five bank groups Foreign banks, Nationalized bank, New private sector banks, Old private sector banks, SBI And Its Associates. Other details of attributes are in appendix table A. Three selected machine learning algorithms Bayesian classification Bayes net local space search algorithm K2, Hill Climb. TAN algorithms are used from- Weka 3.6.2 open source software. For each of these algorithms, five different executions were repeated with different seed values. At each execution the seed value was used to perform stratified cross validation. Thus, the results can be interpreted / analyzed in a more realistic way. In Bayesian, classifier training is very fast. Also the model designed is simple and intuitive. In Bayesian, classifier error is minimized subject to the assumptions of independence of attributes and data satisfying distribution model. The precision results were obtained by Bayesian Classifier Bayes net and Naive Bayes, Bank Group was taken as parent node The Bayesian approach yielded an 73% correctly classified instances and incorrectly classified instances 27%. The main motivation for testing the Bayes net approach was its efficiency, and simplicity, however, the results obtained show that Bayes net TAN local Search structure is effective as desired. Various log score are obtained for different local search algorithms. Bayes Network can be learned using search algorithms and quality measures. Base class for a Bayes Network classifier. Provides data structures (network structure, conditional probability distributions, etc.) and facilities common to Bayes Network learning algorithms like K2.The overall accuracy is another measure commonly used for investigating the quality of classifiers. Accuracy values should be analyzed carefully, because they are not recommended to make decisions about the best classifiers, nonetheless, they can be useful to have intuition about general trends. While precision gives a notion of the proportion of correct predictions out of all the positive predictions, accuracy gives the proportion of correct predictions out of all the examples, either positive or negative. The overall accuracy is calculated using various algorithms and Obtained results are given in Appendix table A. Other statistical measures Confusion matrix, Kappa Statistics, ROC curve True positive rate, false positive rate is displayed in the appendix table B.

TABLE 1: VARIOUS STRUCTURE LEARNING MEASURES

Various Algorithms	TAN	Hill-Climb	K2
Time taken to build model(in seconds)	0.45	0.36	0.14
Correctly Classified Instances	96.50%	72.55%	72.55%
Incorrectly Classified Instances	3.50%	27.45%	27.45%
Kappa statistic	0.9526	0.6363	0.6363
Mean absolute error	0.0153	0.1109	0.1109
Root mean squared error	0.1075	0.3176	0.3176
Relative absolute error	5.17%	37.47%	37.47%
Root relative squared error	27.95%	82.57%	82.57%

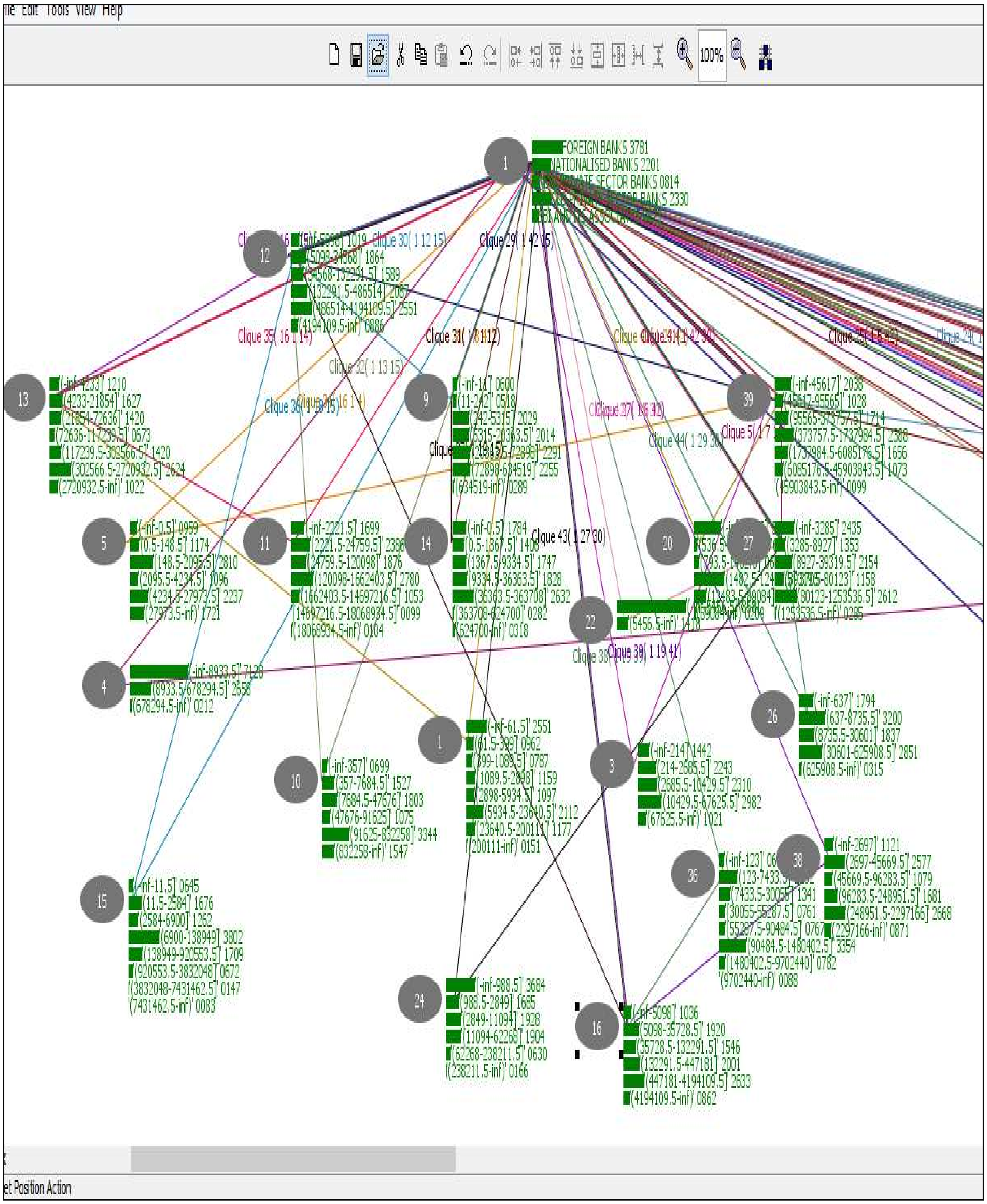
Here TAN algorithm provides the best result in terms of accuracy with 96% correctly classified instances, Kappa statistics 0.9526, Mean squared and Root mean squared error are near to zero.

TABLE 2: BAYESIAN NET LOG SCORE OF VARIOUS ALGORITHM

Various Log Score	TAN	Hill-Climb	K2
Log Score Bayes	-67445.33	-99300.54	-99300.54
Log Score Bdeu	-246987.56	-107581.78	-107581.78
Log Score MDL	-184265.78	-106345.90	-106345.90
Log Score ENTROPY	-97254.47	-99951.50	-99951.50
Log Score AIC	-120373.47	-101650.50	-101650.50

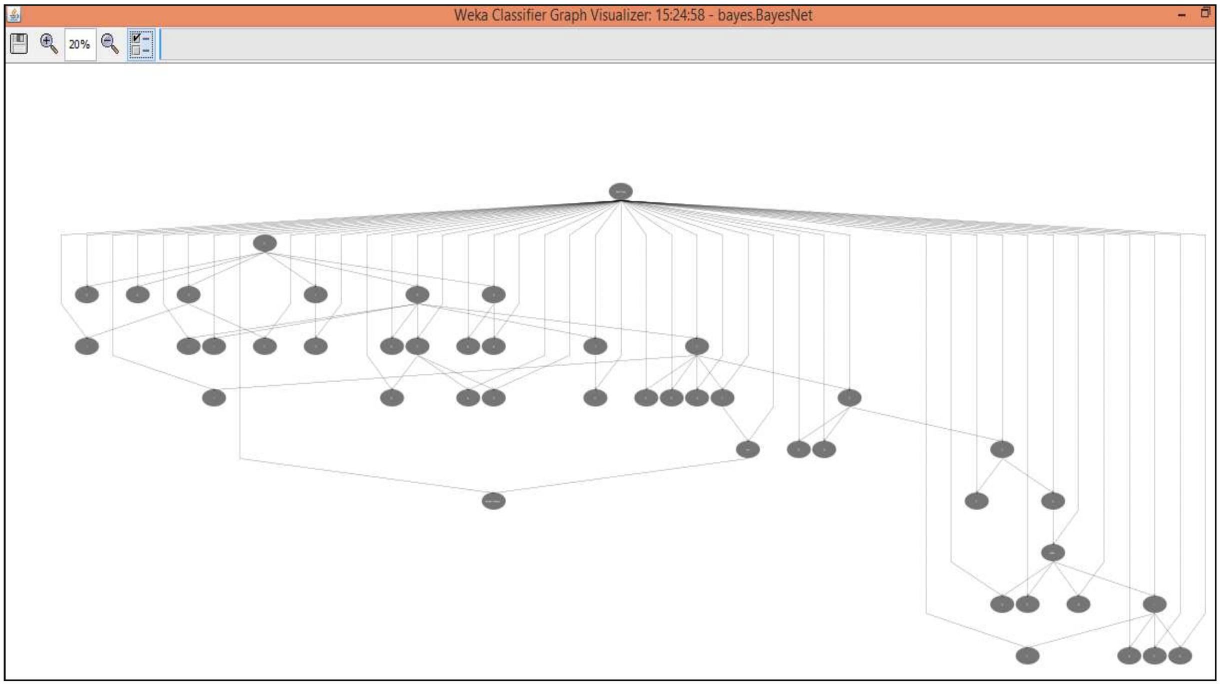
Tan algorithm of structure learning provides best log score Bdeu and second best score MDL compare to Hill climb and K2 algorithm

FIGURE 1: BAYESIAN NET USING BANK GROUP AS PARENT NODE



All 45 attributes were taken and Bayesian Net structure learned with relationship between nodes and probability as parameters obtained for each node.

FIGURE 2: TAN GRAPH OF BAYESIAN NETWORK



Tan graph provides the structure of the graph and dependencies among all 45 attributes.

FIGURE 3: NUMBER OF BANK GROUP

Selected attribute
Name: Bank Group Type: Nominal
Missing: 0 (0%) Distinct: 5 Unique: 0 (0%)

No.	Label	Count
1	FOREIGN BANKS	703
2	NATIONALISED BANKS	409
3	NEW PRIVATE SECTOR BANKS	151
4	OLD PRIVATE SECTOR BANKS	433
5	SBI AND ITS ASSOCIATES	162

Using classification and probabilistic querying we can get the categorical information of various Bank groups information from the data.

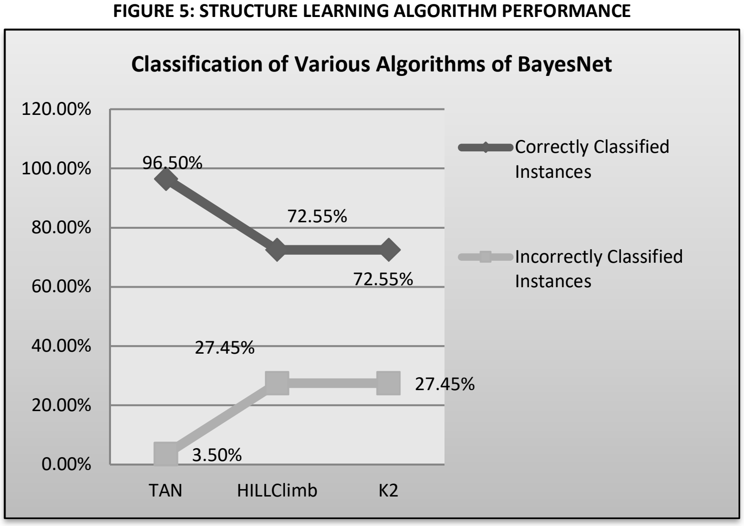
FIGURE 4: PROBABILITY DISTRIBUTION OF BANK GROUP

Probability Distribution Table For Bank Group

FOREIGN BANKS	NATIONALISED...	NEW PRIVATE ...	OLD PRIVATE S...	SBI AND ITS A...
0.378	0.22	0.081	0.233	0.087

The whole database is converted in the form of probability. The above figure display the probability related to bank group data.

FIGURE 5: STRUCTURE LEARNING ALGORITHM PERFORMANCE



Comparison of various measures of classification of various structure learning algorithms

FIGURE 6: PERFORMANCE OF BAYESIAN NET LOG-SCORE

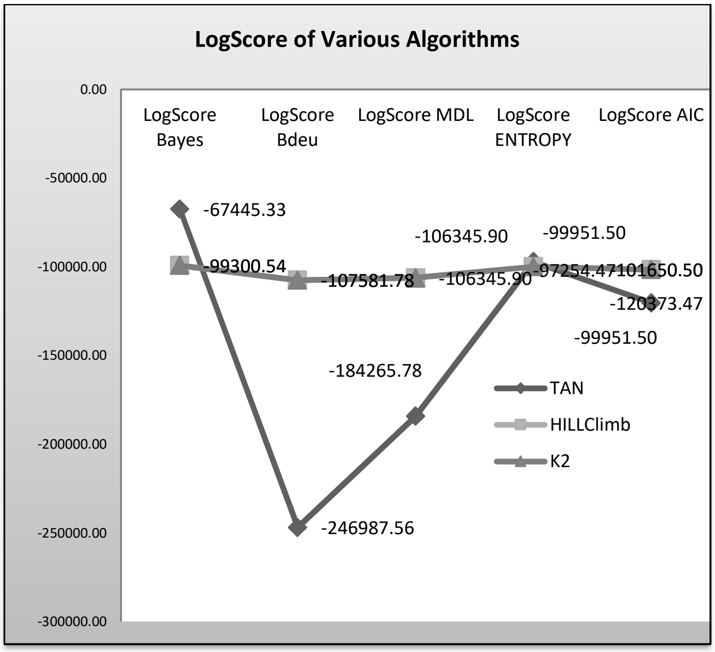


FIGURE 7: MEASURES OF CLASSIFICATION OF BAYESNET

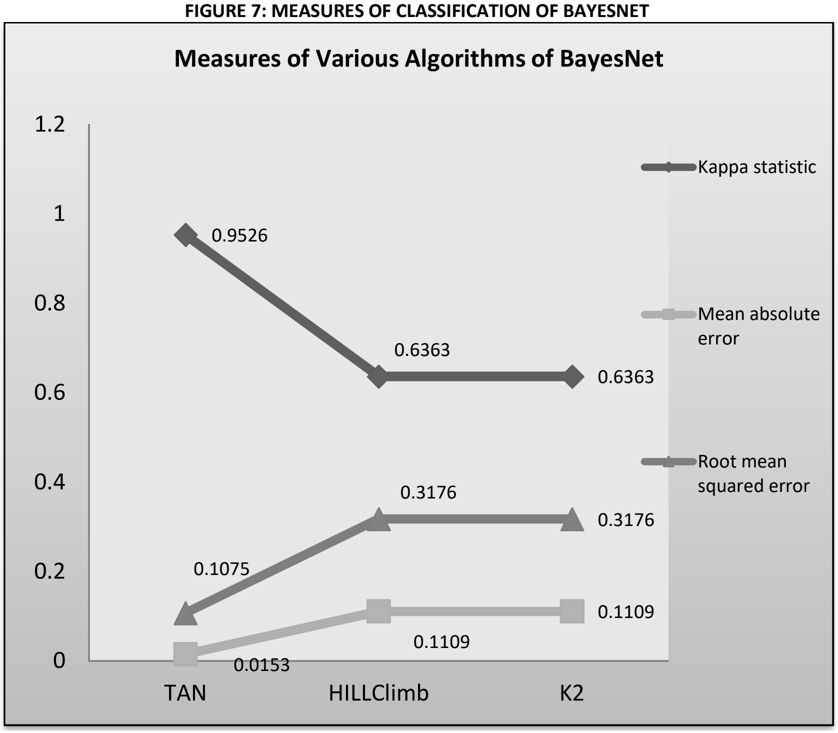
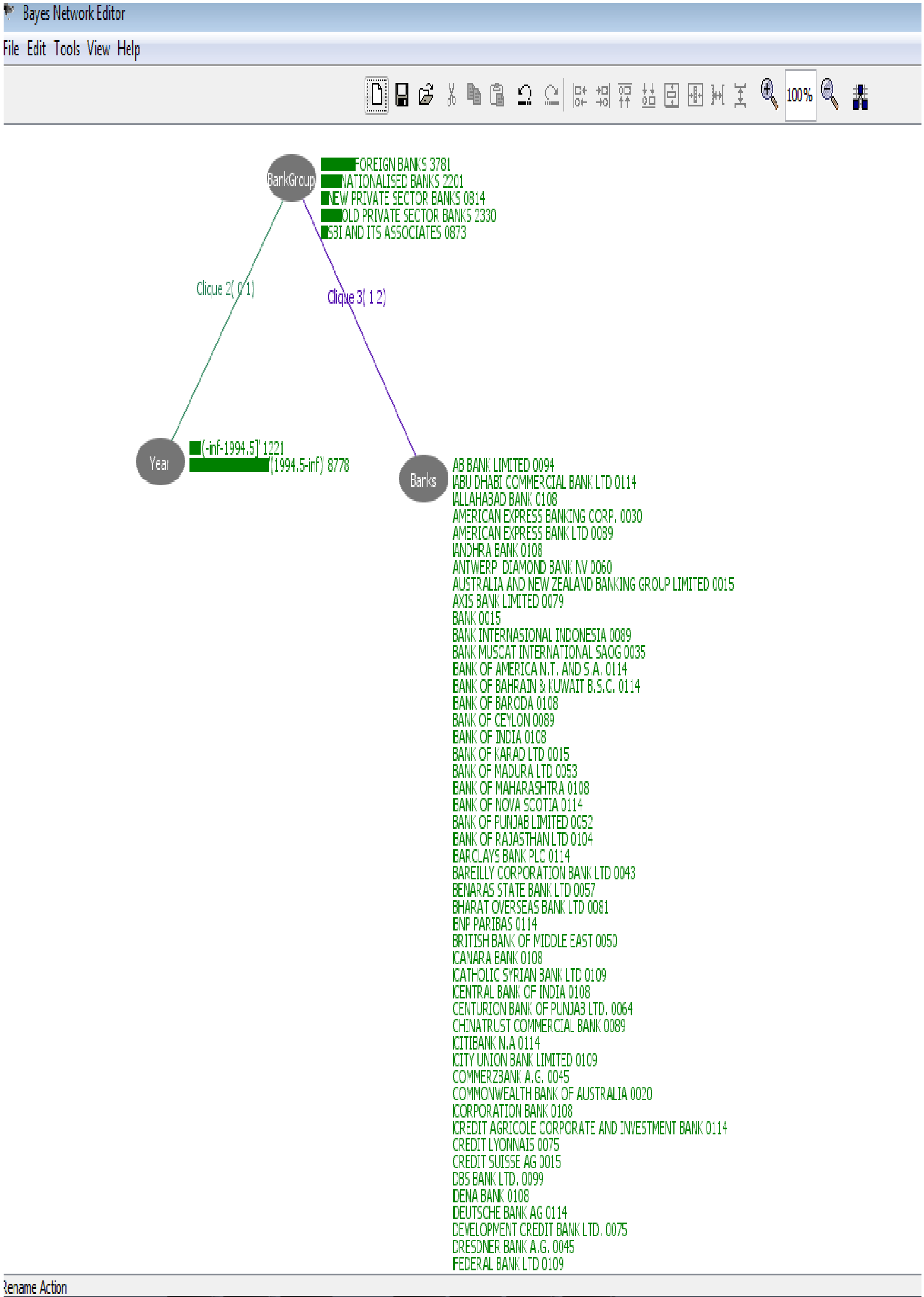


FIGURE 8: LIMITED NODE STRUCTURE - BAYESIAN NET USING BANK GROUP AS PARENT NODE



Only three attribute are taken relationship between nodes and probability obtained for each attribute.

FIGURE 9: NUMBER OF BANKS AND RELATED PROBABILITY OF EACH BANK

Probability Distribution Table For Bank Group

Banks	FOREIGN BANKS	NATIONALISED BANKS	NEW PRIVATE SECTOR BANKS	OLD PRIVATE SECTOR BANKS	SBI AND ITS ASSOCIATES
AB BANK LIMITED	0.897	0.026	0.026	0.026	0.026
ABU DHABI COMMERCIAL BANK LTD	0.915	0.021	0.021	0.021	0.021
ALLAHABAD BANK	0.021	0.915	0.021	0.021	0.021
AMERICAN EXPRESS BANKING CORP.	0.692	0.077	0.077	0.077	0.077
AMERICAN EXPRESS BANK LTD	0.892	0.027	0.027	0.027	0.027
ANDHRA BANK	0.021	0.915	0.021	0.021	0.021
ANTWERP DIAMOND BANK NV	0.84	0.04	0.04	0.04	0.04
AUSTRALIA AND NEW ZEALAND BANKING GROUP LIMITED	0.429	0.143	0.143	0.143	0.143
AXIS BANK LIMITED	0.024	0.024	0.902	0.024	0.024
BANK	0.429	0.143	0.143	0.143	0.143
BANK INTERNASIONAL INDONESIA	0.892	0.027	0.027	0.027	0.027
BANK MUSCAT INTERNATIONAL SAOG	0.733	0.067	0.067	0.067	0.067
BANK OF AMERICA N.T. AND S.A.	0.915	0.021	0.021	0.021	0.021
BANK OF BAHRAIN & KUWAIT B.S.C.	0.915	0.021	0.021	0.021	0.021
BANK OF BARODA	0.021	0.915	0.021	0.021	0.021
BANK OF CEYLON	0.892	0.027	0.027	0.027	0.027
BANK OF INDIA	0.021	0.915	0.021	0.021	0.021
BANK OF KARAD LTD	0.143	0.143	0.143	0.429	0.143
BANK OF MADURA LTD	0.043	0.043	0.043	0.826	0.043
BANK OF MAHARASHTRA	0.021	0.915	0.021	0.021	0.021
BANK OF NOVA SCOTIA	0.915	0.021	0.021	0.021	0.021
BANK OF PUNJAB LIMITED	0.037	0.037	0.852	0.037	0.037
BANK OF RAJASTHAN LTD	0.022	0.022	0.022	0.911	0.022
BARCLAYS BANK PLC	0.915	0.021	0.021	0.021	0.021
BAREILLY CORPORATION BANK LTD	0.053	0.053	0.053	0.789	0.053
BENARAS STATE BANK LTD	0.04	0.04	0.04	0.84	0.04
BHARAT OVERSEAS BANK LTD	0.029	0.029	0.029	0.886	0.029
BNP PARIBAS	0.915	0.021	0.021	0.021	0.021
BRITISH BANK OF MIDDLE EAST	0.81	0.048	0.048	0.048	0.048
CANARA BANK	0.021	0.915	0.021	0.021	0.021
CATHOLIC SYRIAN BANK LTD	0.021	0.021	0.021	0.915	0.021
CENTRAL BANK OF INDIA	0.021	0.915	0.021	0.021	0.021
CENTURION BANK OF PUNJAB LTD.	0.03	0.03	0.879	0.03	0.03
CHINATRUST COMMERCIAL BANK	0.892	0.027	0.027	0.027	0.027
CITIBANK N.A	0.915	0.021	0.021	0.021	0.021
CITY UNION BANK LIMITED	0.021	0.021	0.021	0.915	0.021
COMMERZBANK A.G.	0.789	0.053	0.053	0.053	0.053
COMMONWEALTH BANK OF AUSTRALIA	0.556	0.111	0.111	0.111	0.111
CORPORATION BANK	0.021	0.915	0.021	0.021	0.021
CREDIT AGRICOLE CORPORATE AND INVESTMENT BANK	0.915	0.021	0.021	0.021	0.021
CREDIT LYONNAIS	0.871	0.032	0.032	0.032	0.032
CREDIT SUISSE AG	0.429	0.143	0.143	0.143	0.143
DBS BANK LTD.	0.902	0.024	0.024	0.024	0.024
DENA BANK	0.021	0.915	0.021	0.021	0.021

4. CONCLUSIONS

Various Bayesian classification approaches Bayesnet perform well. The local structure search obtained by Tree Augmented Naive Bayes where the tree is formed by calculating the maximum weight spanning tree using Chow and Liu algorithm perform batter. Graphical models capable of displaying relationships clearly and intuitively with conditional probability of each node. Graphical model can be used to represent indirect in addition to direct causation. For different no of attributes, the Bayes log-score is different. If number of attributes or node increases Bayesian net perform batter. Bayesian networks are directional, thus being capable of representing cause-effect relationships various search algorithm.

5. REFERENCES

1. Agresti A, Categorical Data Analysis Wiley, 2nd edition (2002).
2. Ben-Gal I, Ruggeri F., Faltin F. & Kenett R. Bayesian Networks, Encyclopedia of Statistics in Quality & Reliability, Wiley and Sons. (2007)
3. Changhe Yuan Brandon Malone Learning Optimal Bayesian Networks: A Shortest Path Perspective Journal of Artificial Intelligence Research 48 (2013) 23-65
4. Clement Kirui, Li Hong, Wilson Cheruiyot and Hillary Kirui,Classifiers in Data Mining, IJCSI, International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, March 2013,ISSN, (Print): 1694-0814 | ISSN (Online) 1694-0784
5. David Heckerman and John S. Breese, Casual Independence for probability Assessment and Inference Using Bayesian Networks, IEEE Transactions on Systems, Man and Cybernetics. (1996)
6. Dawn E. Holmes Laxmi C. Jain (Eds.), Studies in Computational Intelligence 156, Innovations in Bayesian Networks, Theory and Applications, Springer, ISBN 978-3-540-850663- 6 (2008)
7. Dr. Bruce Marcot, Dr. Olivier Pourret, (2008) Statistics in practice "Bayesian network a practical guide to applications book published by John Wiley and Sons Ltd.
8. GalitShmueli, Nitin R. Patel and Peter C. Bruce (2010) Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner Second Edition by John Wiley & Sons Ltd
9. Gordon S. Linoff and Michael J.A. Berry (2010) Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Third Edition, ISBN:9780470650936, by John Wiley and Sons Ltd.
10. Heena Timani, Dhaval Panchal, Arzoo Shah, Knowledge discovery from music meta data using semantic web, 2nd International Conference on Emerging Research in Computing Information, Communication and Application ERCICA- Volume 3, Elsevier pp 456-461(2014)

11. Heena TImani, Mayuri Pandya, Data Analysis using probabilistic graphical models, 4th IIMA International Conference on Advance Data Analysis, Business Analytics and Intelligence, Conference Proceeding (2015)
12. Jiawei Han and Micheline Kamber Elsevier publication, Indian Reprint ISBN 978-81- 32-05358 pp 282-327 (2011)
13. Kluwer Academic Publishers, Bayesian Network Classifiers Machine Learning (1997), Manufactured in The Netherlands. 29, 131–1632.
14. M. Nikhita, J. Ashwini, P. Amritha and M.N. Lakshmi, Mining the E-commerce Data to Analyze the Target Customer Behavior International Journal of Industrial Informatics Volume 1, Issue 1, pp-03- 06 (2011)
15. Mark F. Hornic, Erik Marcade, Java Data Mining Strategy, Standard and Practice. A practical Guide for Architecture, Design and Implementation, Morgan Kaufmann Publishers, ISBN 13-978-0-12-370452-8 page 80-1606.
16. Michael Ashcroft, Bayesian networks in business analytics, Proceedings of the Federated Conference on Computer Science and Information Systems ISBN 978-83- 60810-51- 4, pp. 955–961 (2012)
17. Nir Friedman, Dan Geiger, Moises Goldszmidt Bayesian Network Classifiers Machine Learning Kluwer Academic Publishers, Manufactured in The Netherlands. 29, 131–163 (1997)
18. Pearl, J., Fusion, Propagation and structuring in belief networks, Journal. Artificial Intelligence archive. Volume 29 Issue 3, pp 241–288. (1986)
19. Pearl, J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, Book, Publishers Inc. San Francisco, CA, USA ISBN:0-934613- 73-76(1988)
20. Shenoy, Catherine, and Prakash Shenoy, A Bayesian network model of portfolio risk and return. Published in Abu-Mostafa, LeBaron, Lo and Weigend (eds.). Computational Finance, MIT Press, pp 87-106. (2000).

APPENDIX

A.

=== Run information ===
Scheme:weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
Relation: For analysis Assets ver 1
Instances: 1858
Attributes: 45
Year
Bank Group
Banks
Amount in Rupees
1. Cash in hand
2. Balances with RBI
3. Balances with banks in India
4. Money at call and short notice
5. Balances with banks outside India
6.1. Investments in India
6.2. Investments outside India
6. Investments
7A.1. Bills purchased and discounted
7A.2. Cash credits, overdrafts & loans
7A.3. Term loans
7. Advances
7B.1. Secured by tangible assets
7B.2. Covered by Bank/Govt. Guarantees
7B.3. Unsecured
7C.I. Advances in India
7C.II. Advances outside India
8.1. Premises
8.2. Fixed assets under construction
8.3. Other Fixed assets
8. Fixed Assets
9.1. Inter0office adjustments (net)
9.2. Interest accrued
9.3. Tax paid
9.4. Stationery and Stamps
9.5. Others
9. Other Assets
(i) Government securities
(i) Government securities
(iii) Banks
(iii) Others
(iii) Shares
(ii) Other approved securities
(ii) Public sectors
(ii) Subsidiaries and/or joint ventures
(i) Priority sectors
(iv) Debentures and Bonds
(iv) others
Total Assets
(vi) Others
(v) Subsidiaries and/or joint ventures
Test mode:10-fold cross-validation

B.

=== Classifier model (full training set) ===
Bayes Network Classifier
not using ADTree
#attributes=45 #classindex=1
Network structure (nodes followed by parents)
Year(2): Bank Group
Bank Group(5):
Banks(127): Bank Group
Amount in Rupees(1): Bank Group
1. Cash in hand(8): Bank Group
2. Balances with RBI(8): Bank Group
3. Balances with banks in India(5): Bank Group
4. Money at call and short notice(3): Bank Group
5. Balances with banks outside India(6): Bank Group
6.1. Investments in India(9): Bank Group
6.2. Investments outside India(6): Bank Group
6. Investments(7): Bank Group
7A.1. Bills purchased and discounted(7): Bank Group
7A.2. Cash credits, overdrafts & loans(6): Bank Group
7A.3. Term loans(7): Bank Group
7. Advances(6): Bank Group
7B.1. Secured by tangible assets(7): Bank Group
7B.2. Covered by Bank/Govt. Guarantees(7): Bank Group
7B.3. Unsecured(8): Bank Group
7C.I. Advances in India(6): Bank Group
7C.II. Advances outside India(5): Bank Group
8.1. Premises(10): Bank Group
8.2. Fixed assets under construction(2): Bank Group
8.3. Other Fixed assets(6): Bank Group
8. Fixed Assets(8): Bank Group
9.1. Inter0office adjustments (net)(2): Bank Group
9.2. Interest accrued (9): Bank Group
9.3. Tax paid(6): Bank Group
9.4. Stationery and Stamps(9): Bank Group
9.5. Others(5): Bank Group
9. Other Assets(6): Bank Group
(i) Government securities(3): Bank Group
(i) Government securities(7): Bank Group
(iii) Banks(3): Bank Group
(iii) Others(4): Bank Group
(iii) Shares(6): Bank Group
(ii) Other approved securities(8): Bank Group
(ii) Public sectors(8): Bank Group
(ii) Subsidiaries and/or joint ventures(4): Bank Group
(i) Priority sectors (8): Bank Group
(iv) Debentures and Bonds(6): Bank Group
(iv) others(6): Bank Group
Total Assets(7): Bank Group
(vi) Others(4): Bank Group
(v) Subsidiaries and/or joint ventures(5): Bank Group
=== Stratified cross-validation ===
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
0.74 0.072 0.862 0.74 0.796 0.939 FOREIGN BANKS
0.853 0.077 0.759 0.853 0.803 0.968 NATIONALISED BANKS
0.536 0.06 0.443 0.536 0.485 0.874 NEW PRIVATE SECTOR BANKS
0.651 0.082 0.707 0.651 0.678 0.899 OLD PRIVATE SECTOR BANKS
0.716 0.057 0.545 0.716 0.619 0.937 SBI AND ITS ASSOCIATES
0.6992 0.073 0.741 0.726 0.729 0.931 Weighted Avg.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue, as well as on the journal as a whole, on our e-mail **infoijrcm@gmail.com** for further improvements in the interest of research.

If you have any queries, please feel free to contact us on our e-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward to an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-
Co-ordinator

DISCLAIMER

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals