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.,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Polandwith IC Value of 5.09 &number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 2401 Cities in 155 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

Weinvite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Education, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; 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; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor 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; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email address: infoijrcm@gmail.com.

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/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '_____' for possible publication in your journals.

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

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mentionthe following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of 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 required to be below **500 KB**.
e) Abstract alone 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 and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **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 stops at the end.

6. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES:** These should be simple, crystal clear, centered, separately numbered &self explained, and **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.

11. **EQUATIONS:**These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **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 are supposed to follow **Harvard Style of Referencing**. 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 italics. 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 parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

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, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, KurukshetraUniversity, 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 Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

AN EFFICIENT MINING PROCEDURE FOR GENE SELECTION BY USING SELECT ATTRIBUTES

S.ANUSUYA
STUDENT
DEPARTMENT OF CSE
BHARATH UNIVERSITY
CHENNAI

R.KARTHIKEYAN
ASST. PROFESSOR
DEPARTMENT OF CSE
BHARATH UNIVERSITY
CHENNAI

ABSTRACT

We propose a novel gene selection algorithm based on select attributes. The proposed algorithm eliminates gene redundancy automatically and yields a very small number of cancer related genes. Using the selected genes on the cancer classification, the robust classification accuracy has been produced across different types of classification methods. By comparing the number of selected genes and the classification accuracy obtained by our methods with others, our gene selection algorithm is very competitive compared to most recent gene selection methods. In addition, it has been convinced by the literature research that our identified genes are biologically relevant to cancer. Therefore our method will be a useful supplementary tool for the future studies in the application of microarray datasets.

KEYWORDS
Classification, DNA Microarray, Gene Selection, Select attribute, WEKA.

1. INTRODUCTION

The main reason for gene mining is to find and separate genes that are characterized for conferring essential traits. The availability of molecular biological technique is more. It has allowed for rapid development and identification of nucleic acid derived sequence.

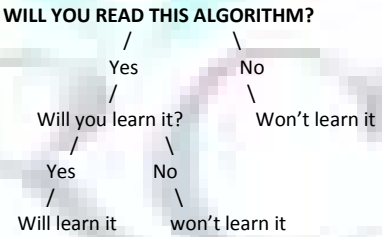
WHERE IS GENE LOCATED*?*

A cell is made up of chromosomes. 23 pairs of chromosomes present in a single cell. The Chromosome is made up of genes. The genes consist of DNA which is made up of four chemical letters .i.e A, C, T, G. In a cell both DNA and RNA are present. Both did a replication process.DNA was convert into RNA is called transcription process. RNA was convert into DNA is called reverse transcription process. Protein was obtained from the RNA with the help of translation process. The advantages of gene mining is, it is used in clonning method,pregnancy cases,agriculture. The disadvantage of gene mining is,high tech,costly. In this paper large number of attributes are present,so using the select attributes and then classify with various algorithm it produce different output for various algorithms.compare all output and find which one is low output, it is the final accuracy. Furthermore, using the selected genes on the cancer classification, the robust classification accuracy has been produced by some different classification methods.

2. METHODS USED FOR GENE MINING

2.1 **CLASSIFICATION**

Classification is a data mining algorithm that creates a step-by-step guide for how to determine the output of a new data instance. The tree it creates is exactly that: a tree where by each node in the tree represents a spot where a decision must be made based on the input, and you move to the next node and the next until you reach a leaf that tells you the predicted output. Sounds confusing, but it's really quite straightforward. Let's look at an example.



This simple classification tree seeks to answer the question "Will you understand classification trees algorithm?" At each node, you answer the question and move on that branch, until you reach a leaf that answers yes or no. This model can be used for any unknown data instance, and you are able to predict whether this unknown data instance will learn classification trees by asking them only two simple questions. That's seemingly the big advantage of a classification tree it doesn't require a lot of information about the data to create a tree that could be very accurate and very informative.

2.2. **GAUSSIAN PROCESS**

Implements Gaussian process for regression without hyper parameter tuning. To make choosing an appropriate Noise level easier, this implementation applies normalization / standardization to the target attribute as well. Missing values are replaced by the global mean / mode. Nominal attributes are converted to binary ones.

2.3. **LINEAR REGRESSION**

Regression is the easiest technique to use, but is also probably the least powerful.. This model can be as easy as one input variable and one output variable.Of course, it can get more complex than that, including dozens of input variables. In effect, regression models all fit the same general pattern. There are a number of independent variables, which, when taken together, produce a result a dependent variable. The regression model is then used to predict the result of an unknown dependent variable, given the values of the independent variables.

2.4. **MULTILAYERED PERCEPTRON**

It is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a direct graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron with a non linear activation function. MLP

utilizes a supervised learning technique called back propagation for training the network. MLP is a modification of the standard linear perceptron and can distinguish data that is not linearly separable.

2.5. **SMO REG**

SMO reg implements the support vector machine for regression. The parameters can be learned using various algorithms. The algorithms is selected by setting the reg optimizer. The most popular algorithm(reg smo improved) is due to shevade, keerthi et al and this is the default reg optimizer.

3. NUMERICAL EXPERIMENTS

In order to exam our algorithm, we carry out some experiments on datasets, namely Breast cancer datasets. Breast cancer dataset have an attribute is 11 and instance is 699. Using select attributes, select few attributes out of 11 attributes and classify with various algorithm. Find the output. Do the procedure repeatly until the output becomes low. This process is called as selecting best method by various search models. Use all methods and find best method.

3.1 **Find accuracy**

Breast cancer datasets

Instance 699

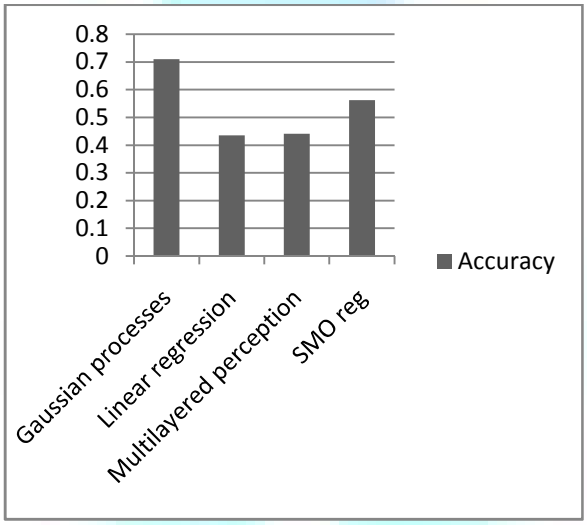
Attributes 11

There are Sample code number, Clump thickness, Uniformity of cell size, Uniformity of cell shape, Marginal adhesion, Single epithelial cell size, Bare nuclei, Bland chromatin, Normal nucleoli, Mitoses, Class. In select attributes, cross validation is select in the attributes select mode. Then number of folds are represent in %. Here choose only four attributes there are clump thickness, bare nuclei, bland chromatin, class.

TABLE 1: SELECT ATTRIBUTES 4

Algorithm	Accuracy
Gaussian processes	0.7105
Linear regression	0.4353
Multilayered perception	0.4408
SMO reg	0.562

FIG. 1: GRAPH FOR 4 ATTRIBUTES

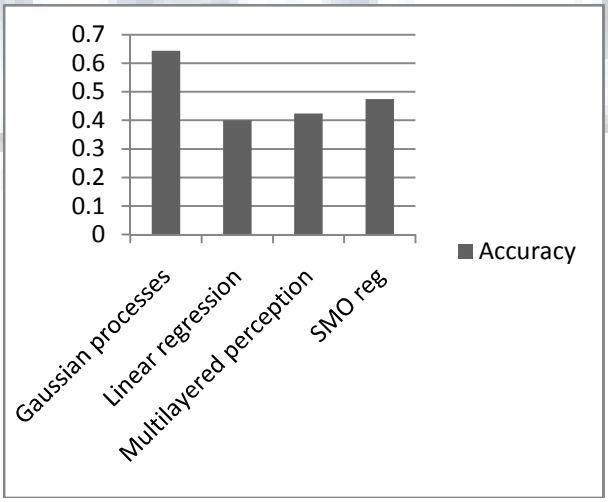


Here choose only five attributes. They are clump thickness, Uniformity of cell size, Bare nuclei, Bland chromatin, Class.

TABLE 2: SELECT ATTRIBUTES 5

Algorithm	Accuracy
Gaussian processes	0.6444
Linear regression	0.4003
Multilayered perception	0.4243
SMO reg	0.4747

FIG. 2: GRAPH FOR 5 ATTRIBUTES

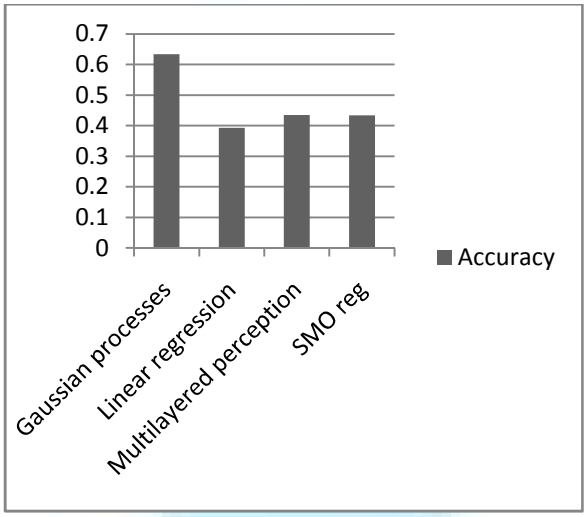


Here choose only six attributes. They are Clump thickness, Uniformity of cell size, Bare nuclei, Bland chromatin, Normal nucleoli, Class.

TABLE 3: SELECT ATTRIBUTES 6

Algorithm	Accuracy
Gaussian processes	0.6342
Linear regression	0.3933
Multilayered perception	0.4351
SMO reg	0.4335

FIG. 3: GRAPH FOR 6 ATTRIBUTES

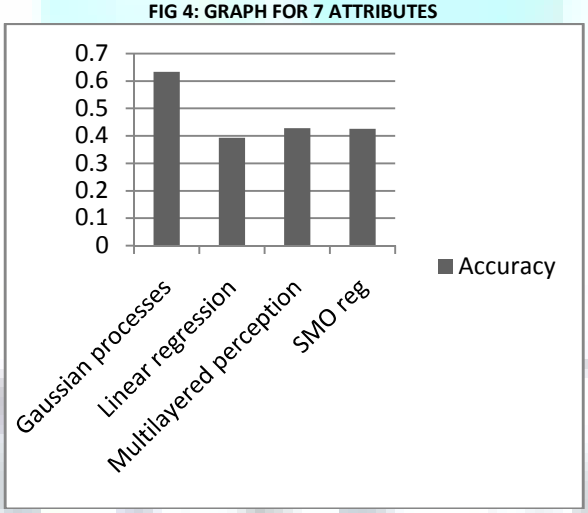


Here choose only seven attributes. They are clump thickness, uniformity of cell size, uniformity of cell shape, bare nuclei, bland chromatin, normal nucleoli, class.

TABLE 4: SELECT ATTRIBUTES 7

Algorithm	Accuracy
Gaussian processes	0.6338
Linear regression	0.3929
Multilayered perception	0.4284
SMO reg	0.4262

FIG 4: GRAPH FOR 7 ATTRIBUTES



4. COMPARISON WITH PREVIOUS WORK

The gene selection problem and classification problem have been studied by several authors. For comparison, we list the number of selected genes and classification accuracy obtained by different methods and it is listed above. Here select few attributes and find the accuracy. When the accuracy becomes low, this process is stopped.

5. BIOLOGICAL INTERPRETATION OF THE SELECTED GENES

In this section, explain about the genes in this datasets selected by the algorithm and analyze their relationship with the occurrence of disease, which are based on the identified genes closely correlated with the pathogenesis of specific or general diseases.

In summary, the majority of the genes selected by our method are relevant to the pathogenesis of disease. Some of them have definite biological meaning while the others remain to be explored. It reflects the complexity of cancerous pathogenesis.

6. CONCLUSION

Gene selection procedure is conducted on the microarray datasets through these algorithms. It can reduce the gene redundancy automatically and collect a very small number of related genes. By using the selected genes on the classification, rather high classification accuracy is obtained. The number of selected genes

and the classification accuracy obtained by our methods was compared with others, our gene selection algorithms are very hard or difficult compared to most recent gene selection methods.

7. ACKNOWLEDGMENT

The authors would like to thank the management of Bharath university for their support and department of computer science and engineering encouragement for related works.

8. REFERENCES

1. A Novel Gene Selection Algorithm Based on Binary Decision Tree, 1Zenghui Yang, 1bZhixia Yang, 2Yongcui Wang.1, College of Mathematics and Systems Science, Xinjiang University, 830046, Urumqi, China,
2. Alizadeh A, Eisen M, Eric DR , "Distinct types of diffuse large B-cell lymphoma identified by gene expression Profiling", Nature vol.403,pp. 503--511, 2000.
3. Gulub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, BloomTeld CD, Lander ES, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", Science vol.286,pp. 531--537, 1999.
4. Ma S, Huang J, "Regularized ROC method for disease classification and biomarker selection with microarray data", Bioinformatics vol.21,pp. 4356--4362, 2005.
5. Marchiori E, Sebag M, "Bayesian learning with local support vector machines for cancer classification with gene expression data", Lect Notes Comput Sci vol.3449,pp. 74--83, 2005.
6. Nguyen DV, Rocke DM, "Tumor classification by partial least squares using microarray gene expression data", Bioinformatics vol.18,pp. 39--50, 2002.
7. Peng Y, Li W, Liu Y, "A hybrid approach for biomarker discovery from microarray gene expression data for cancer classification", Cancer Informatics vol.2,pp. 301--311, 2006
8. Wang Y, Tetkoa IV, Hallb MA, Frankb E, Faciusa A, Mayera KF, Mewesa HW, Gene selection from microarray data for cancer classification-- a machine learning approach, Comput Biol Chem vol.29,pp. 37--46, 2005.
9. Yang ZX, Zhao YM, "A Hybrid Approach for the Multi-class Gene Selection Problem ", JCIT: Journal of Convergence Information Technology, Vol. 6, No. 5, pp. 1 -17, 2011
10. Yu L, Liu H, "Redundancy based feature selection for microarray data", In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, USA, pp. 737--742, 2004.

REQUEST FOR FEEDBACK

Dear Readers

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

I would like to request you tosupply 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 youhave 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 an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

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