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We invite unpublished novel, original, empirical and high quality research work pertaining to the recent developments & practices in the areas of Computer Science & Applications;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

IMPROVING CLASSIFICATION PERFORMANCE USING ENSEMBLE LEARNING APPROACH

JYOTSANA GOYAL
STUDENT
DEPARTMENT OF COMPUTER SCIENCE
LAKSHMI NARAIN COLLEGE OF TECHNOLOGY
RGPV UNIVERSITY
INDORE

Er. AMIT VAJPAYEE
ASST. PROFESSOR
LAKSHMI NARAIN COLLEGE OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE
RGPV UNIVERSITY
INDORE

ABSTRACT

The data mining techniques are used for evaluation of the data in order to find and represent the data in such manner by which the applications are becomes beneficial. Therefore, different kinds of computational algorithms and modeling's are incorporated for analyzing the data. These computational algorithms are help to understand the data patterns and their application utility. The data mining algorithms supports supervised as well as unsupervised techniques of data analysis. This work is aimed to investigate about the supervised learning technique specifically performance improvements on classification techniques. The proposed classification model includes the multiple classifiers namely Bayesian classifier, k-nearest neighbor and the c4.5 decision tree algorithm. By nature of the outcomes and the modeling of the data these algorithms are functioning differently from each other. Thus, a weight based classification technique is introduced in this work. The weight is a combination of outcomes provided by the implemented three classifiers in terms of their predicted class labels. Using the weighted outcomes, the final class label for the input data instance is decided. The implementation of the proposed working model is performed with the help of JAVA and WEKA classes. The results obtained by experimentation of the proposed approach with the vehicle data set demonstrate the high accurate classification results. Thus, the proposed model is an effective classification technique as compared to single model implementation for classification task.

KEYWORDS

data mining, classification, supervised learning, ensemble learning, performance improvement.

INTRODUCTION

Data mining is a task of data analysis. In this process the similarity and difference among the available set of data is computed. Using this technique, the patterns are established that is used for different applications in order to make decisions, classifications and predictions. According to the algorithm applied on data these applications are performed. These various kinds of algorithms which are categorized in two major categories namely supervised approaches and unsupervised approaches. In this work the supervised learning approaches are studied more specifically the classification algorithms. Basically, the classification algorithms first learn from the data and then used for classifications of similar patterns.

The learning of the algorithm depends upon the quality of data and the present noise in the data patterns. Additionally, the process of learning is also creating impact of learning. But in supervised learning techniques it is expected to generate the high accurate results in less time resource consumption. Therefore, the performance improvement of classification technique is one of the essential tasks in data mining in terms of classification accuracy and the resource consumption. The performance can be improved in three major phases by enhancing the quality of data therefore the feature extraction techniques are used, or by manipulating the learning algorithm in these techniques the traditional algorithms are modified for improving the learning capability or by implementing the ensemble learning algorithms. In this work the study of ensemble learning is the primary aim of the work additionally enhancing the traditional approach of learning is the second key aim.

PROPOSED WORK

This chapter provides the details about the proposed system design and their functional aspects. Therefore, first the overview of the system is provided and then the system model is described. Finally, the proposed approach is summarized using the algorithm steps.

A. SYSTEM OVERVIEW

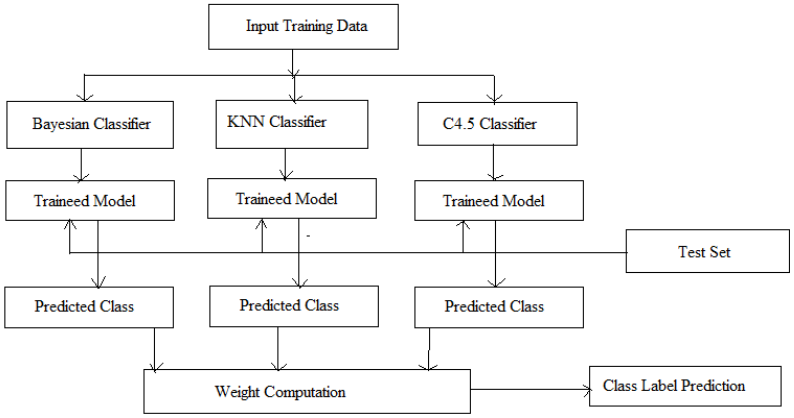
In data mining techniques, the performance improvements can be performed in all the three possible stages of data analysis. In data mining the three phases are dataset input, training using the algorithm and the testing of the algorithm. During the input of data, the data can also be transformed for recovering the essential features for training, or by pre-processing techniques the quality of data can be optimized. In next the modification and enhancements can also be performed by optimizing the learning algorithms. Finally, in the post processing stages the performance can also be optimized. These post processing techniques are termed as the ensemble learning techniques such as bagging and boosting.

By the motivation of the ensemble learning techniques the proposed work is intended to design a new weighted technique for improving the classification performance. This technique incorporates more than on learner in classification system and using the predicted outcomes of these learners the combined outcome for the single instance of data is generated. The combination of results is given using the weight building. The weights are computed on the basis of some intermediate coefficients. That is user dependent values which are vary between 0-1. Additionally, according to the requirements these values can be adjusted by the designer. By using these coefficients and the predicted outcomes of the classifiers the final outcome of classifier is predicted. The proposed technique's overview is described in this section and the next section explains the working of the proposed methodology.

B. SYSTEM DESIGN

The proposed classification technique is demonstrated using figure. The different components of the system are also listed in this diagram. The detailed description of the involved components is given as:

FIGURE 1: PROPOSED SYSTEM



Input dataset: the data mining systems are works on the data for finding the valuable patterns. In this experiment the structured data is used for analysis and design. Therefore, the vehicle dataset is used for experimentation and design. This dataset is available in ARFF (attribute relationship file format) and CSV (comma separated file) in UCI repository. Not only can the vehicle data set function in this system other ARFF files can also works with the system.

Bayesian classifier: The standard approach to Bayesian classification uses the chain rule to decompose the joint distribution:

$$\Pr(C, A_1, A_2, \ldots, A_k) = \Pr(C)\Pr(A_1, A_2, \ldots, A_k|C) \ldots \ldots \ldots (1)$$

The first term on the right-hand side of (1) is the prior probability of the class labels. These can be directly estimated from the training data, or from a larger sample of the population. For example, we can often get statistics on the number of, say, breast cancer occurrences in the general population. The second term on the right-hand side of (1) is the distribution of attribute values given the class label. The estimation of this term is usually more complex, and we elaborate on it below.

Once we have an estimate of $\Pr(C)$ and $\Pr(A_1, A_2, \ldots, A_k|C)$ we can use Bayes rule to get the conditional probability of the class given the attributes:

$$\Pr(C|A_1, A_2, \ldots, A_k) = \alpha \Pr(C)\Pr(A_1, A_2, \ldots, A_k|C) \ldots \ldots \ldots \ldots (2)$$

where α is a normalization factor that ensures that the conditional probability of all possible class labels sums up to 1. (In practice, we do not need to explicitly evaluate this factor because it is constant for a given instance.) Using (2) we can classify new instances by combining the prior probability of each class with the probability of the given attribute values given that class.

The Naive Bayes classification algorithmic rule is a probabilistic classifier. It is based on probability models that incorporate robust independence assumptions. The independence assumptions usually don't have an effect on reality. So, they're thought of as naive. You can derive probability models by using Bayes' theorem (proposed by Thomas Bayes). Based on the nature of the probability model, you'll train the Naive Bayes algorithm program in a very supervised learning setting. In straightforward terms, a naive Bayes classifier assumes that the value of a specific feature is unrelated to the presence or absence of the other feature, given the category variable. There are two types of probability as follows:

- Posterior Probability [P (H/X)]
- Prior Probability [P (H)]

Where, X is data tuple and H is some hypothesis. According to Bayes' Theorem

$$P\left(\frac{H}{X}\right) = \frac{P\left(\frac{X}{H}\right)P(H)}{P(X)}$$

KNN classifier: The K-nearest-neighbor algorithm measures the distance between a query scenario and a set of scenarios in the data base. The distance between these two scenarios is estimated using a distance function d (x, y), where x, y are scenarios developed through features, like

$$X = \{x_1, x_2, x_3, \ldots\}$$

$$Y = \{y_1, y_2, y_3, \ldots\}$$

The frequently used distance functions are absolute distance measuring using:

$$d_A(x, y) = \sum_{i=1}^{N} |x_i - y_i|$$

And second is Euclidean distance measuring with:

$$d_A(x, y) = \sum_{i=1}^{N} \sqrt{x_i^2 - y_i^2}$$

The overall KNN algorithm is running in the following steps:

1. Store the output values of the M nearest neighbors to query scenario Q in vector r = {r_1,……,r_m} by repeating the following loop M times:
   a. Go to the next scenario S_i in the data set, where I is the current iteration within the domain {1……P}
   b. If Q is not set or q < d (q, S_i): q ←d (q, S_i), t ←O_i
   c. Loop until we reach the end of the data set.
   d. Store q into vector c and t into vector r.
2. Calculate the arithmetic mean output across r as follows:

$$\bar{r} = \frac{1}{M}\sum_{i=1}^{M} r_i$$

3. Return r as the output value for the query scenario q

C4.5 classifier: C4.5 (developed by Quinlan, 1993) an algorithm that learns the decision-tree classifiers, it has been observed that C4.5 performs short in the domain where there is pre-entrance of continuous attributes compared with the learning tasks with mostly separate attributes. For instance, a system which looks for well-defined decision tree with 2 levels and then put comments [8]:

"The accuracy of trees made with T2 is equalized or even exceed trees of C4.5 upon 8 out of all the datasets, with the entire except one that have incessant attributes only."

INPUT: An exploratory data set of data (D) portrayed with the means of discrete variables.

OUTPUT: A decision tree say T which is constructed by means of passing investigational data sets.

1) A node (X) is created;
2) Check if the instance falls in the same class.
3) Make node (X) as the leaf node and assign a label CLASS C;
4) IF attribute list is empty, THEN
5) Make node(X) a leaf node and assign a label of most customary CLASS;
6) Now choose an attribute which has highest information gain from the provided attribute List, and then marked as the $test_{attribute}$;
7) Confirming X in the role of the $test_{attribute}$;
8) In order to have a recognized value for every $test_{attribute}$ for dividing the samples;
9) Generating a fresh twig of tree that is suitable for $test_{attribute}$ = att_i from node X;
10) Take an assumption that Bi is a group of $test_{attribute}$ =att_i in the samples;
11) If Bi = = NULL, THEN
12) Add a new leaf node, with label of most common class;
13) ELSE a leaf node is going to be added and returned by $Generate_{decisionTree}$.

This section provides the implementation of the classical C4.5 algorithm the next section provides the concept for enhancing the model performance.

Trained model: after processing of the training set the system generates the mathematical model for evaluation of similar kinds of data patterns. In this context, the Bayesian classifier computes the probability for events, KNN classifier computes the distance among the available data instances and the C4.5 algorithm generates the tree structure for the input training set. This tree structure help to invoke the test set attributes for predicting the class labels.

Test set: the test set is additional sample data which is used for evaluation of data model. After preparing the data models the test set instances are evaluated one by one to generate the class labels. In this context, the KNN algorithm finds the likely instances of data with their classes, similarly the Bayesian classifier computes the posterior probability for finding the class labels and finally the C4.5 generate the class labels by using the tree invocation process.

Predicted class: all the implemented classification techniques accept the test data as input and for each instance of the data the classifiers generates the class labels. According to the generated classes and the data instance is produced in next step for computing the weights and generating the final classes for the instance data.

Weight computation: the predicted class labels by the trained classifiers are used here for computing the final class label for data instance. In order to compute the weights the class labels are treated here in terms of numerical values. In order to compute the weights for predicted outcome the following formula is used:

$$W = C_1 * w_1 + C_2 * w_2 + C_3 * w_3$$

Where, W is the computed weight of the class, and C_1, C_2 *and* C_3 are the classes predicted by the classifiers. And w_1, w_2 *and* w_3 are the coefficient selected between 0-1 and such that $w_1 + w_2 + w_3 = 1$. That is user dependent function which can be regulated according to the priority of the designer. In this presented work the weight coefficients are equally distributed for 0.33 to all the classifiers.

Class labels: that is the final outcome of the proposed ensemble learning based classification technique. That is the most likely outcome which is predicted by the implemented classifiers on the basis of weights.

C. Proposed Algorithm

This section provides the understanding about the proposed classification algorithm. That algorithm is summary of the entire steps involved in the proposed working model. Following table contains the proposed algorithm steps:

TABLE 1: PROPOSED ALGORITHM

Input: training dataset T, testing dataset Ts Output: class labels of test dataset C
Process: 1. $R = ReadTrainingSet(T)$ 2. $B_{model} = Bays.Train(R)$ 3. $K_{model} = KNN.Train(R)$ 4. $C4.5_{model} = C45.Train(R)$ 5. $Ts_m = readTestingDataset(Ts)$ 6. $for(i = 1; i \leq m, i++)$ a. $C_1 = B_{model}.classify(Ts_i)$ b. $C_2 = K_{model}.classify(Ts_i)$ c. $C_3 = C4.5_{model}.classify(Ts_i)$ d. $W = 0.33 * C_1 + 0.33 * C_2 + 0.33 * C_3$ e. $C_i = Math.Round(W)$ 7. End for 8. Return C

RESULTS ANALYSIS

The performance of the proposed hybrid algorithm over different classifier algorithms are evaluated and compared in this chapter. Therefore, the different performances factors are evaluated and demonstrated in this section i.e. accuracy, error rate, memory consumption and time complexity.

A. Accuracy

The performance of the classifier in terms of accuracy is given in this section. The performance evaluation of proposed hybrid classifier is evaluated using implementation of all three classifiers. The accuracy of the system can be given using the following formula.

$$\text{Accuracy} = \frac{\text{Total Correctly Classified Samples}}{\text{Total Samples Avilable}} X100$$

The performance of the proposed system and another individual classifier algorithm is provided using the following figure and table. In all the experimentation, we take a similar data set for k-NN, C4.5, Bayesian and proposed hybrid classification approach of different experiments. The blue bars in the given figure contain the hybrid classification performance for the original datasets without any change. The brown bar of the figure demonstrates C4.5; similarly, orange and yellow bar depicts the kNN and Bayesian classifiers. According to the obtained results the proposed classifier performs much better learning for proposed data quality enhancement and improvement classification of the system as compared to other classifiers. In this graph, proposed classifier obtaining higher accuracy rate for large number of correct data classification.

GRAPH 1: COMPARE ACCURACY

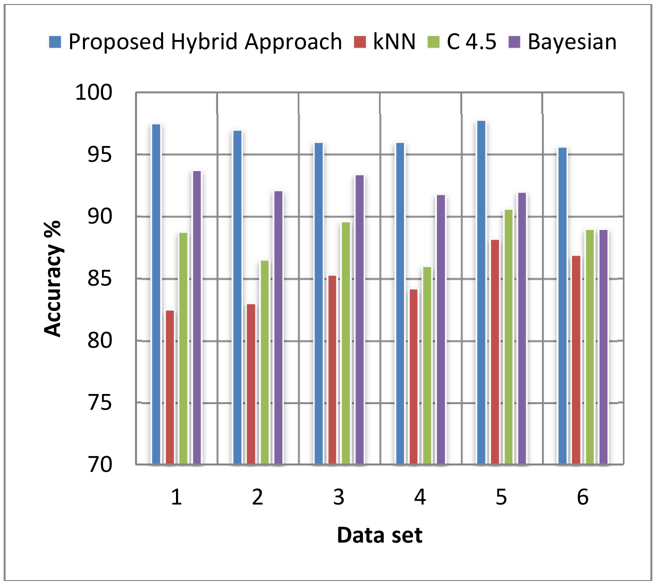


TABLE 2: TABULAR FORM OF ACCURACY

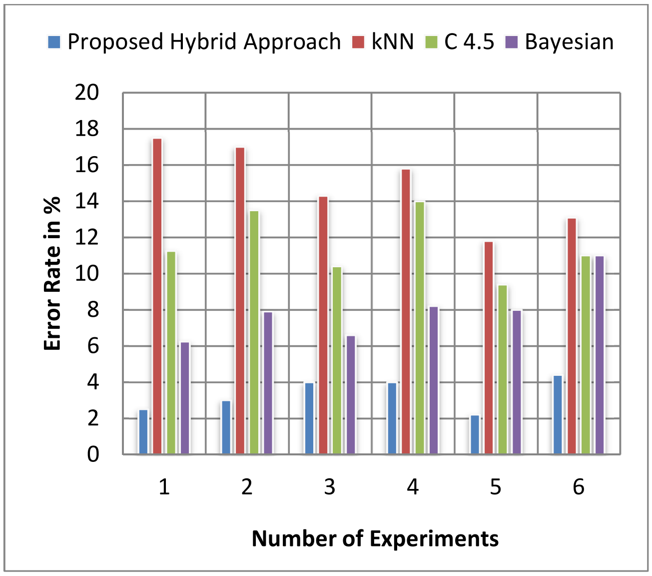
Number of Experiments	Proposed Hybrid Approach	k-NN	C 4.5	Bayesian
1	97.5	82.5	88.75	93.75
2	97	83	86.5	92.1
3	96	85.3	89.6	93.4
4	96	84.2	86	91.8
5	97.8	88.2	90.6	92
6	95.6	86.9	89	89

B. Error Rate

The error rate of the classifier reports the amount of data that are not properly recognized during the classification. The error rate of the classifiers can be evaluated using the following formula.

$$\text{Error Rate} = \frac{\text{Misclassified Samples}}{\text{Total Samples to Classify}} X100$$

GRAPH 2: COMPARE ERROR RATE



In this figure and table shows the error rate of the implemented classification system. Therefore, to represent the performance of the algorithms the X axis contains the different experimental individually and the Y axis shows the error rate percentage of all classifiers. In above demonstration, high percentage of error rate is producing by kNN classifier which has high number of misclassified pattern, whereas other two classifiers have average error rate performance. According to the obtained performance the proposed technique produces the less error rate where less number of data have been misclassified as compared to the other remain 3 classifiers. Therefore, the proposed technique is much effective for classification improvement

TABLE 3: TABULAR FORM OF ERROR RATE

Number of Experiments	Proposed Hybrid Approach	k-NN	C 4.5	Bayesian
1	2.5	17.5	11.25	6.25
2	3	17	13.5	7.9
3	4	14.3	10.4	6.6
4	4	15.8	14	8.2
5	2.2	11.8	9.4	8
6	4.4	13.1	11	11

C. Memory Consumption

Memory consumption of the system also termed as the space complexity in terms of algorithm performance. This can be calculated using the following formula:

$$\text{Memory Consumption} = \text{Total Memory} - \text{Free Memory}$$

The amount of memory consumption depends on the amount of data reside in the main memory, therefore that effect the computational cost of an algorithm execution. The performance of the implemented proposed classifier along with other traditional classifier for data classification is given using figure and table. For reporting the performance, the X axis of figure contains the different number of code execution and the Y axis shows the respective memory consumption of proposed and other classifiers during the execution of system in terms of kilobytes (KB). According to the obtained results the performance of the proposed Hybrid Classification approach consuming moderate space of the system when data being executing. By the given graph, C4.5 and Bayesian are taking little space for execution where proposed approach and K-NN taking high number of space. Therefore, we conclude that it is the limitation of the project of space complexity.

GRAPH 3: COMPARE MEMORY CONSUMPTION

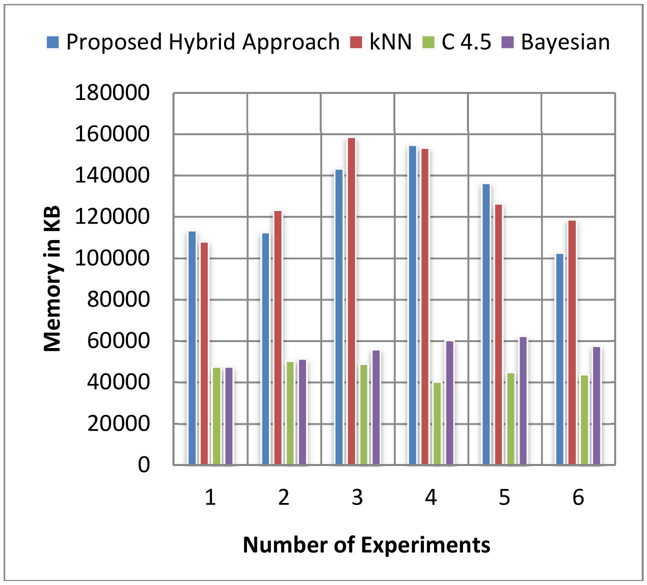


TABLE 4: TABULAR FORM OF MEMORY

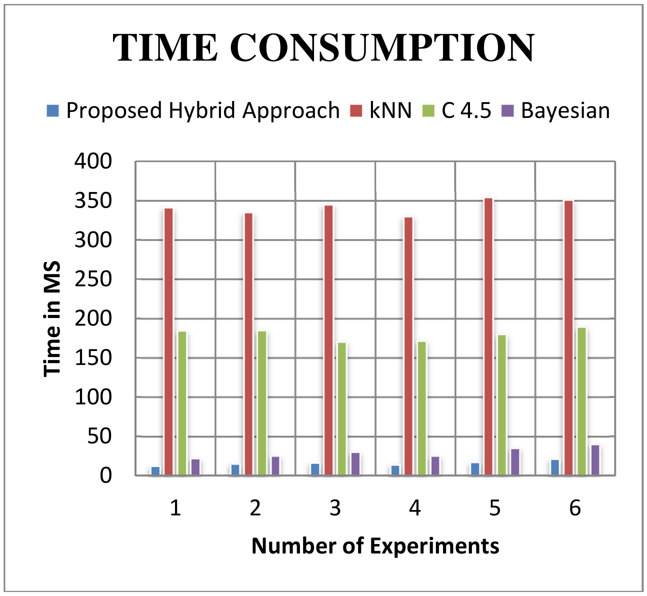
Number of Experiments	Proposed Hybrid Approach	k-NN	C 4.5	Bayesian
1	113335	108030	47511	47525
2	112547	123251	50321	51246
3	143261	158456	48792	55846
4	154812	153251	40215	60251
5	136251	126254	44782	62147
6	102542	118541	43625	57593

D. Time Consumption

The amount of time required to classify the entire test data is known as the time consumption. That can be computed using the following formula:

$$\text{Time Consumed} = \text{End Time} - \text{Start Time}$$

GRAPH 4: COMPARE TIME CONSUMPTION



The time consumption of the proposed algorithm is given using figure and table. In this diagram the X axis contains the experimental scenario and the Y axis contains time consumed in terms of seconds for implemented classifiers. According to the comparative results analysis and given the performance hybrid classification approach is executing small amount of time to process the dataset. In similar way, Bayesian also taking small amount of time but in respective of proposed it is little high amount of time consuming. We analysis the resulting graph where different data set produces result monotonically as consume less time to process dataset.

TABLE 5: TABULAR FORM OF TIME VALUES

Number of Experiments	Proposed Hybrid Approach	k-NN	C 4.5	Bayesian
1	12	341	184	22
2	15	335	185	25
3	16	345	170	30
4	14	330	171	25
5	17	354	180	35
6	21	351	189	40

Using the Template

The main of the proposed work to investigate and design an ensemble learning based technique is accomplished successfully. This chapter provides the conclusion of the conducted research work and the obtained experimental results. In addition of that the feasible future work is also suggested in this chapter.

CONCLUSION

Data mining techniques supports various kinds of data modeling according to the requirements of the application and required outcomes. In addition of that the utilization of algorithms are also depends on the nature and type of data. In this presented work, the classification techniques are main aim of study. The classification techniques are the supervised learning approaches that are used for prediction, pattern recognition, decision making and other complicated task. In these applications, the high accurate outcomes are required to be obtained. Therefore, different techniques of performance optimization of classifiers are utilized. In this presented work, the post processing of classifiers is studied and according to the obtained conclusion a new model for learning and accurate classification is presented. The proposed technique is motivated form the ensemble learning approach where the goodness of multiple classifiers is combined to optimize the final classification outcomes.

Thus, the proposed work is intended to improve the performance of classification by using the traditional classifiers. Therefore, the ensemble learning approach is proposed for study. In this context three base classifiers namely Bayesian classifier, KNN (k-nearest neighbor) and C4.5 decision tree is selected for designing and developing the proposed system. Basically, the Bayesian classifier is a probabilistic approach of classification, kNN is the distance based classification approach and the C4.5 algorithms is a rule based data model that works on the basis of tree data modeling. Therefore, the functional process of the system is different from each other and need to find a suitable manner by which the outcomes of the classification can be optimized accordingly. Initially all the classifiers are trained using the similar training samples. During training the actual functionality of the classifiers are remain same. In further a test sample is also prepared that is the 30% of entire training data which is randomly selected from the initial training set. The test set is applied on the trained models and their predicted outcomes are gathered. These outcomes are used in next process for computing the final class labels for the data instances.

The implementation of the proposed system requires the JAVA technology and the WEKA class libraries to be implemented for finding required results. During different number of experiments, the obtained results are noticed and their mean values are reported in the below table.

TABLE 6: MEAN PERFORMANCE

Parameters	Proposed	K-NN	C4.5	Bays
Accuracy	96.65 %	85.09 %	88.41 %	92.01 %
Error	3.35 %	14.91 %	11.59 %	7.99 %
Time	15.83 MS	342.66 MS	179.83 MS	29.5 MS
Memory	127124.66 KB	131297.16 KB	45874.33 KB	55768 KB

According to the obtained performance as given in above table the proposed technique is able to generate more accurate classification as compared to single classification technique implementation. Thus, the proposed technique is acceptable where the accuracy is required as compared to resource preservation.

FUTURE WORK

According to the computed results and the performance the proposed technique is found suitable and efficient for utilizing in the other applications where the accurate decisions are required. In near future, the following is work is feasible for extending the current implemented system.

1. In near future, the work is extended for rule based classification enhancement and their optimization

2. The work can also be extendable with more classifiers such as neural network, SVM and others
3. Need to improve the required memory consumption for utilizing in the other applications.

REFERENCES

1. Parvin, Hamid, Miresmaeil MirnabiBaboli, and Hamid Alinejad-Rokny, "Proposing a classifier ensemble framework based on classifier selection and decision tree", Engineering Applications of Artificial Intelligence 37 (2015): pp. 34-42.
2. Introduction to Data Mining and Knowledge Discovery, Dunham, M. H., Sridhar, S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, 1st Edition, 2006.
3. Phridvi Raj MSB., GuruRao CV (2013) Data mining – past, present and future – a typical survey on data streams. INTER-ENG Procedia Technology 12, pp. 255 – 263
4. Veepu Uppal and Gunjan Chindwani, "An Empirical Study of Application of Data Mining Techniques in Library System", International Journal of Computer Applications (IJCA), Volume 74– No.11, July 2013.
5. Shalev-Shwartz, Shai, and Shai Ben-David. Understanding machine learning: From theory to algorithms, Cambridge University press, 2014.
6. Nilsson, Nils J. "Introduction to machine learning, An early draft of a proposed textbook." (1996).
7. Kotsiantis S.B. Supervised machine learning: a review of classification techniques. Informatica 31:249–268, 2007.
8. Duda, R. O., Hart, P. E., and Stork, D. G. Pattern Classification, Wiley-Interscience, 2nd edition, (2001).
9. Ghahramani, Z. Unsupervised learning. In O. Bousquet, G. Raetsch, & U. von Luxburg (Eds.), Advanced lectures on machine learning. Berlin: Springer-Verlag (2004).
10. Meenakshi and Geetika, "Survey on Classification Methods using WEKA", International Journal of Computer Applications, Vol. 86, No.18, January 2014.

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