IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

Content Based Alternate Medicine Recommendation by Using Random Forest Algorithm

Ankita D. Rewade[1*], Sudhir W. Mohod[2], Sharad P. Bargat[3]

[1,2]Dept. of Computer Science Engineering, Bapurao Deshmukh College of Engineering, Nagpur University, Sevagram, Wardha, India
[3]Dept. of Mechanical Engineering, Terna Engineering College, Mumbai University, Nerul, Navi Mumbai, India

[*]*Corresponding Author: asbargat@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i4.11631168 | Available online at: www.ijcseonline.org

Accepted: 21/Apr/2019, Published: 30/Apr/2019

Abstract – This paper explores the application of the random forest algorithm for the alternate medicine recommendation. In this work the users able to search alternate medicine for their particular prescribed medicine. The main aim of the proposed system is to provide the users with alternate recommendation of medicines based on their content and shows the medicines as per ascending order of cost. It also has a facility to users to provide ratings for a particular medicine. Proposed System Architecture for content-based systems, one needs to assess the similarity of any many distinct medicines. If there exists a sufficient textual description for each medicine, this can be achieved via random forest algorithm. However, a user's input medicine must contain all contain of medicine otherwise no similarity measuring can be performed.

Keywords **-** Cost and Content Based Recommendation, Random Forest, Healthcare, Alternate Medicine.

I. INTRODUCTION

Health information is one of the most widely concerned topics on the Web. A survey in 2013 by the Pew Internet and American Life Project found that 59% of adults have looked online for health topics, and with 35% of respondents focusing on diagnosing a medical condition online [1]. Behind the data, we find that more and more people are caring about the health and medical diagnosis problem. However, there are still many people losing their lives due to medication errors. According to the administration's report, more than 200 thousand people in China, even 100 thousand in USA, die each year due to medication errors [2]. More than 42% medication errors are caused by doctors because experts write the prescription according to their experiences which are quite limited [3]. There are some facts that may lead to these issues: (i) many hospitals lack either doctors or medical experts for critical illness, (ii) expert diagnosis is mainly depended on the expert's experience, especially for those inexperienced novices, which are hard to avoid mistakes. Meanwhile, most diagnosis case data in hospitals is still kept untouched and has not been used for mining, so that the value behind the data cannot be explored.

Alternate Medicine System is beneficial to both service providers and users. They reduce transaction costs of finding and selecting items in an online shopping environment. Recommendation systems have also proved to improve decision making process and quality. In e-commerce setting, Alternate Medicine System enhances revenues, for the fact that they are effective means of selling more products.

Medicines are prescribed by the doctor based on the drug content of the medicine used for curing the disease. To properly cure disease patients always look around for same drug which may be very expensive or even unavailable at the time of requirement. Due to this patients have to pay much higher price for same product. The scope of our system is to provide alternatives for medicines when the prescribed medicine is either too costly or unavailable at the time of requirement helping people get their treatment on time and at a budget friendly way. Existing Systems can't provide proper recommendation to user for medicines as they are based on collaborative filtering approach which requires data from other users as well. Also the medicine recommendation must have higher accuracy as it is very important for saving lives of people.

II. PROPOSED RANDOM FOREST ALGORITHM

RF (Medicine med)
Step 1: Read medname in M
Step 2: Connect to DB using
getcon()

Step 3: Fetch M details from DB
Step 4: If resultset = NULL
goto EXIT
Step 5: Fetch content into CT[ ] array
Step 6: FOR I = 1: N CT DO
COMPARE CONTENT WITH OTHER MEDICINES
Step 7: IF FOUND
INCREMENT CT[ ]
ELSE
goto STEP 6
Step 8: SORT CT[ ] based on price
Step 9: RETURN CT[ ] with alternates
Step 10: EXIT

In this way the proposed work carried out using the proposed algorithmic strategy. Many text classifiers have been proposed in the literature using machine learning techniques, probabilistic models, etc. They often differ in the approach adopted: decision trees, naıve-Bayes, rule induction, neural networks, nearest neighbours, and lately, support vector machines.

III. PROCESSING STEPS OF PROPOSED SYSTEM ARCHITECTURE

- **Dataset Generation**

Medicine dataset needs to be generated for multiple medicines having same ratio which can be provided as an alternative to each other. Also for cost analysis it is necessary to know the market cost for such medicines. This dataset will be created for a total of 100 medicines. Figure 1 shows the screen shot of dataset generation.

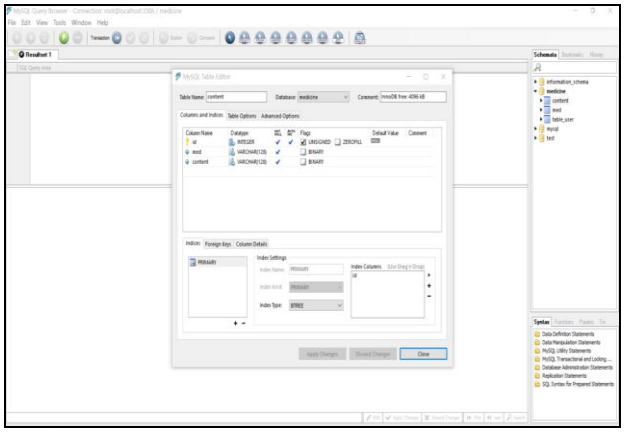

Figure 1: Dataset Generation

- **Content Matching**

Content matching is done for all the contents fetch from the name of medicine input by user. It is then classified with the help of Random Forest algorithm.

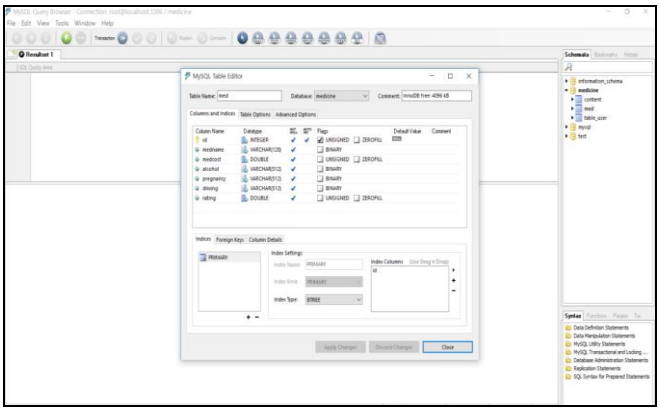

Figure 2: Content Matching

- **Dataset Pre-processing**

As dataset gets generated it is necessary to pre-process it for any null values if provided and the data should be cleaned and stored into Database for further processing

- **Data Clustering based on contents and costing**

Data clustering needs to be done for grouping similar medicines based on their contents and also it is required for cost based analysis as well. Screenshot shows the 0 and 1 condition to identify the User or Admin is going to login.

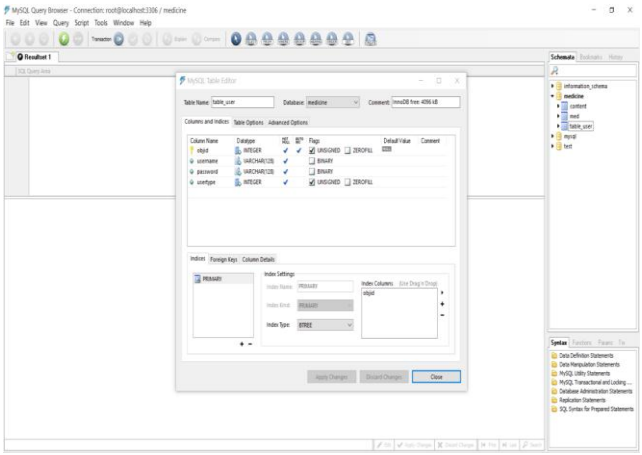

Figure 3: Identification of User or Admin

- **Medicine classification and recommendation**

At last ones the user inputs some medicine there is a requirement of finding alternate medicines for users which can only be done using classification of input medicines using some classification algorithm.

IV. MECHANISM OF SORTING

- Figure 4 explored a basic example of working of random forest for proposed system.
- Based on multiple decision tree fused in single graph, we have placed salt composition for fused graph.

- In second stage content (salt composition) are fused for another checking. In last stage we get alternates in destination nodes.

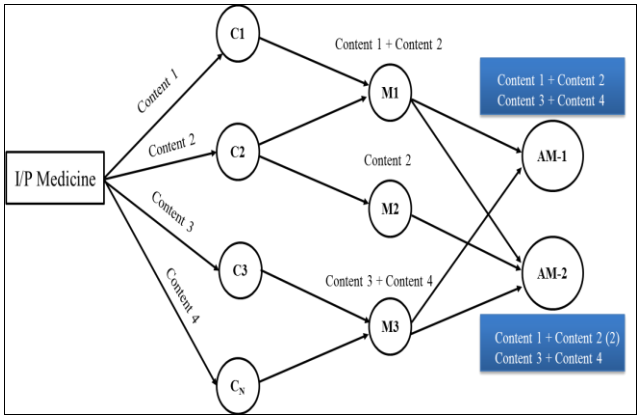

Figure 4: Working of Algorithm with Medicine

V. EXAMPLE WITH LOSAR-H TABLET

Figure 5 shows the actual alternative medicine recommendation process.

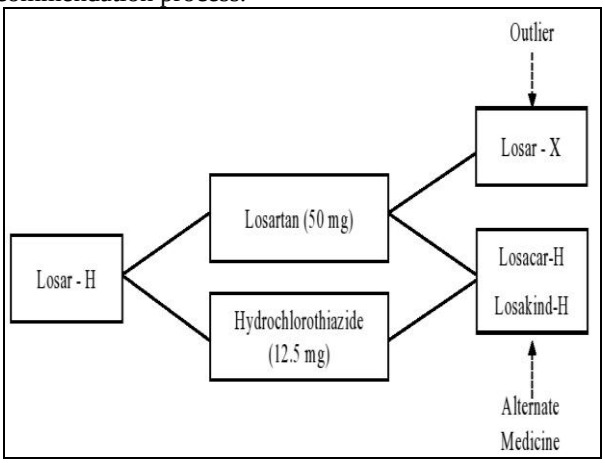

Figure 5: Example with Single Tablet

In Figure 5, the example of Losar-H table has been taken to search alternative medicine. Losar-H table consisting of two main content namely 'Losartan (50mg)' and 'Hydrochlorothiazide (12.5mg)'. Losar X is an Outlier we can say because it consisting only one content i.e 'Losartan (50mg)'. Finally system show two alternatives which are 'Losacar-H Tablet' and 'Losakind-H Tablet'. In this way the alternative medicine recommendation works

VI. ADVANTAGES OF PROPOSED SYSTEM

- Fast availability of medicines
- Lower cost so cost efficient
- Alternative option availability
- Better Health care

VII. PROPOSED SYSTEM FLOW CHART

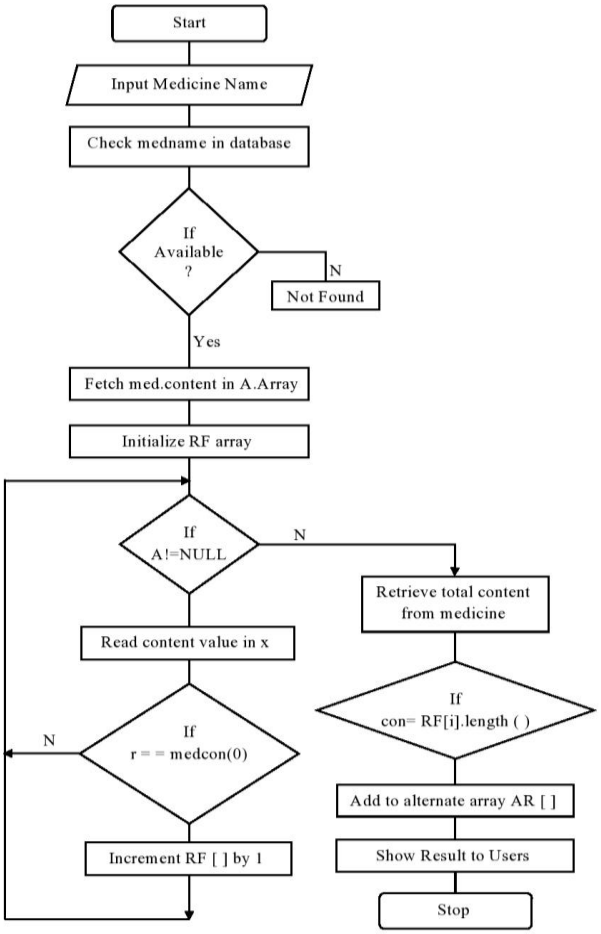

Figure 6: Proposed System Flow Chart

VIII. SYSTEM EXECUTION OF PROPOSED SYSTEM

- **Home Page**

The execution of system is carried in following manner. Figure 7 shows the first appearance of the proposed system. This is the required home page of Proposed System. Both Admin and User can login with the same. The User will have to first register himself.

Figure 7: Home Page

➢ **Login Page**

Figure 8 shows the login page has to manage the details of Username, Password, Email, Changing Password, Permission and Authorization. It manages all the information about Username, medicine, Permission and Authorization. Login page is common for both the users and Admin. By putting unique Id and password Admin and User will be able to operate the System.

Figure 8: Login Page

➢ **Add Medicine Page**

Figure 9 shows the 'Add Medicine' form. Admin can update medicine by inserting medicine details. Medicines directly get stored in the database.

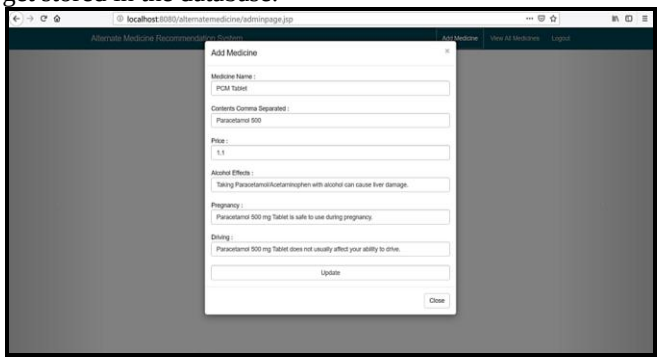

Figure 9: Add Medicine Page

IX. TIME COMPARISON ANALYSIS

The time comparison analysis has been done to analyse the performance of k-NN, Decision tree and Random Forest for the different size of data set. The data set is arranged in the size of 25 medicines, 50 medicines, 75 medicines and 100 medicines with their corresponding time in sec.

Table 1: Time Comparison Evaluation Result

Number of Medicines	Time in Second		
	k-NN	**Decision Tree**	**Random Forest**
25 Medicines	1.23	1.9	0.89
50 Medicines	2.5	2.95	1.1
75 Medicines	2.85	3.2	1.15
100 Medicines	4.5	4.3	1.9

Table 2 shows the time required for classification and recommendation of alternate medicine using of k-NN, Decision tree and random forest algorithms for the 25 medicines, 50 medicines, 75 medicines and 100 medicines size of data set.

The following graph shows the time required for classification and recommendation of alternate medicine using of k-NN, Decision tree and random forest algorithms for the 25 medicines, 50 medicines, 75 medicines and 100 medicines size of data set.

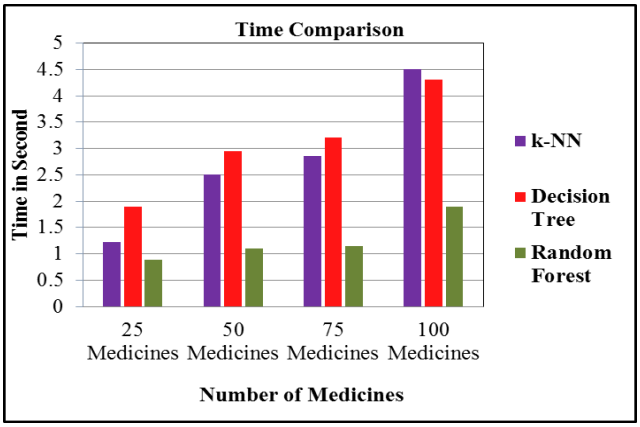

Figure 10: Time Comparison Graph

From the graphical analysis it is noted that the Random Forest provides fastest classification for the large data set as compared to k-NN and Decision tree. This is a result of capability of random forest algorithm to handle thousands of input variables with efficient running without variable deletion. Hence it is highly accurate classifier for large size of data set.

X. ACCURACY ANALYSIS

There are many ways to evaluate the skill of a prediction model. An approach in the related field of information retrieval (finding documents based on queries) measures precision and recall.

The accuracy analysis has been done to analyse the performance of k-NN, Decision tree and Random Forest for the precision and recall factors. Table 6.3 shows the time in second for k-NN, Decision tree and Random Forest algorithm.

When making a prediction for a binary or two-class classification problem, there are two types of errors that one could make.

➢ **False Positive:** Predict an event when there was no event.

- **False Negative**: Predict no event when in fact there was an event.

Precision is a ratio of the number of true positives divided by the sum of the true positives and false positives.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad \text{i.e}$$

Precision = TP / TP + FP

Precision describes how good a model is at predicting the positive class. Precision is referred to as the positive predictive value.

Recall is calculated as the ratio of the number of true positives divided by the sum of the true positives and the false negatives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad \text{i.e.}$$

$$\text{Recall} = \frac{TP}{TP+FP+TN+FN}$$

Total Accuracy = Precision × 100

Table 3: Precision and Recall Comparative

Factors	Time in Second		
	k-NN	**Decision Tree**	**Random Forest**
Precision	0.66	0.71	0.97
Recall	0.45	0.48	0.51

The follwing graph shows the time required for the precision and recall for k-NN, Decision tree and random forest algorithms for the 25 medicines, 50 medicines, 75 medicines and 100 medicines size of data set.

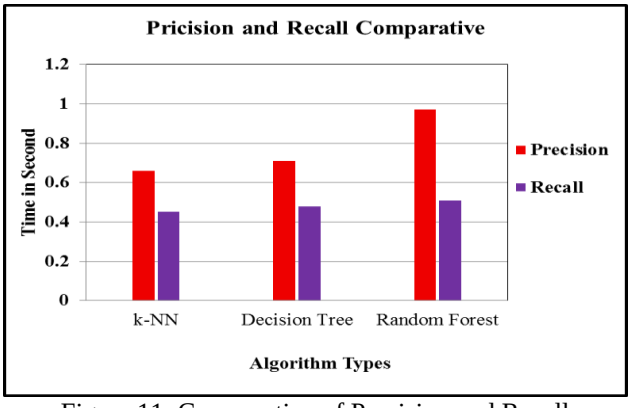

Figure 11: Comparative of Precision and Recall

Figure 10 shows that the Random Forest provides highest precision and lower recall time for the given data set. This is because of Random Forest built by using decision trees, which are sensitive to the distribution of the classes. This shows the efficient running of Random Forest algorithm than k-NN and Decision tree algorithm for the given dataset. Hence it is highly accurate classifier for large size of data set.

XI. APPLICATIONS

- Mobile friendly web application design for alternate medicine recommendation.
- Cost based filtering application for rating costing purpose.
- Medicines side effect checker for allergic patients
- The random forest algorithm is used in a lot of different fields, like Banking, Stock Market, Medicine and E-Commerce.
- Random forests are widely used in large data bases where need to handle thousands of input variables without variable deletion and efficient running is required
- It is applicable where most accurate learning algorithms is required
- It is used to produces a highly accurate classifier for many data sets

XII. CONCLUSION

Proposed system provides the alternate medicine information to the user which is less costly as compared to prescribed medicines and with its effect in different condition of the users. At the same time user can also see the rating given by the other users and also able to provide rating for a particular medicine. Random Forest is fully parallelizable to go even faster and ability to handle data without pre-processing. In Random Forest Data does not need to be rescaled, transformed, or modified. It gives the best alternate predictor selection from large number of medicine.

This work explores the strong application of data mining in the field of medical recommendation systems. In this project dataset over 100 plus medicines from different companies and brands has used to do alternate medicine recommendation based on content of medicine and then filter it based on rating and cost based analysis. Through experimental results found that more than 95% of the medicines have a lower cost based alternative available with a higher rating. The random forest algorithm has been used for classification of medicines based on the costing and rating of the medicines provided by users.

ACKNOWLEDGMENT

I would like to extend my gratitude and sincere thanks to my Guide and Head of department **Dr. S. W. Mohod** for his constant guidance and valuable suggestions during the course of my work.

I also would like extend my gratitude and sincere thanks to **Dr. G. V. Thakre, Principal,** BDCE, Sevagram, Wardha for providing necessary facilities and support during the work.
I also extend my sincere thanks to **Prof. S. P. Bargat**, Assistant Professor in Mechanical Engineering Department, Terna Engineering College, Nerul, Navi Mumbai for his constant motivation and support during the course of my work.

REFERENCES

[1] Naoki Shino, Ryosuke Yamanishi, Junichi Fukmoto, *"Recommendation System For Alternative-ingredients Based On Co-occurrence Relation On Recipe Database And The Ingredient Category",* 5th International Congress On Advanced Applied Informatics, **IEEE**, **2016**.

[2] Weiwei Lin, Ziming Wu, Longxin Lin, Angzhan Wen, And Jin Li, *"An Ensemble Random Forest Algorithm For Insurance Big Data Analysis"*, Recent Advances In Computational Intelligence Paradigms For Security And Privacy For Fog And Mobile Edge Computing, **IEEE**, **5th July 2017**.

[3] Leslie Wilson, Fatema A. Turkistani, Wei Huang, Dang M. Tran, Tracy Kuo Lin, *"The Impact Of Alternative Pricing Methods For Drugs In California Workers' Compensation System: Fee-schedule Pricing",* **May 2018**.

[4] Torgyn Shaikhinaa, Dave Loweb, Sunil Dagad, David Briggsc, Robert Higginse, Natasha Khovanovaa, *"Decision Tree And Random Forest Models For Outcome Prediction Inantibody Incompatible Kidney Transplantation",* Biomedical Signal Processing and Control, **Elsevier, 2017**.

[5] Shyong (Tony) K. Lam John Riedl. *"Shilling Recommender Systems for Fun and Profit",* Group Lens Research Computer Science and Engineering University of Minnesota Minneapolis, **2004**.

[6] Bradley N. Miller, *"Toward a Personal Recommender System",* Workshop: Beyond Personalization 2005 IUI'05, San Diego, California, USA, **9th January, 2005**.

[7] Deepti Sisodiaa, Dilip Singh Sisodiab, *"Prediction of Diabetes using Classification Algorithms"* Procedia Computer Science, **Published by Elsevier**, Vol. **132** pp.**1578-1585, 2018.**

[8] Braun L, Tiralongo E, Wilkinson J, Spitzer O, Bailey M, Poole S, Dooley M., *"Perceptions, use and attitudes of pharmacy customers on complementary medicines and pharmacy practice"e".* BMC Complement Alternate Medicine **2010**.

[9] MacLennan A., Myers S., Taylor A., "*The continuing use of complementary and alternative medicine in South Australia: costs and beliefs in 2004",* Med J Aust., Vol.**184:5**, **2006**.

[10] Xue CCL, Zhang AL, Lin V., Da Costa C., Story DF, *"Complementary and alternative medicine use in Australia: a national population-based survey".* Journal of Alternate Complement Med. 13: pp. **643-650, 2007.**

[11] Braun L "*Integrative Pharmacy: threat or fantasy*", J Compl Med., 8: pp. **41-42**, **2009**.

[12] Semple S. J., Hotham E., Rao D., Martin K., Smith C. A., Bloustien G. F. "*Community pharmacists in Australia: barriers to information provision on complementary and alternative medicines".* Pharm World Sci. 28: pp. **366-373**, **2006**.

[13] J. Pradeep Kandhasamy, S. Balamurali, *"Performance Analysis of Classifier Models to Predict Diabetes Mellitus",* Procedia Computer Science, **Published by Elsevier**, pp. **45-51**, **2015**.

[14] Soudabeh Khodambashi, Alexander Perry, Oystein Nytro, *"Comparing User Experiences on the Search-based and Content-based Recommendation Ranking on Stroke Clinical Guidelines - A Case Study"* Procedia Computer Science, **Published by Elsevier,** pp. **260-267**, **2015**.

[15] Kourosh Modarresi, *"Recommendation System Based on Complete Personalization"* Procedia Computer Science, **Published by Elsevier, Vol. 80,** pp. **2190-2204, 2016**.

[16] Rubal, Dinesh Kumar, *"Evolving Differential evolution method with random forest for prediction of Air Pollution" "* Procedia Computer Science, **Published by Elsevier,** Vol. **132,** pp. **824–833, 2018**.

[17] Mo Hai, You Zhangc, Yuejin Zhanga *"A Performance Evaluation of Classification Algorithms for Big Data"* Procedia Computer Science, **Published by Elsevier Vol.122** pp. **1100-1107, 2017**.

[18] Charu Kathuria, Deepti Mehrotra, Navnit Kumar Misra, *"Predicting the protein structure using random forest approach"* Procedia Computer Science, **Published by Elsevier Vol.132,** pp. **1654–1662, 2018**.

[19] Meenu Shukla, Sanjiv Sharma *"Analysis of Efficient Classification Algorithm for Detection of Phishing Site"* International Journal of Scientific Research in Computer Science and Engineering, Vol.**5**, Issue.**3**, pp.**136-141**, **June 2017.**

[20] J.V.N. Lakshmi, Ananthi *Sheshasaayee "A Big Data Analytical Approach for Analyzing Temperature Dataset using Machine Learning Techniques"* International Journal of Scientific Research in Computer Science and Engineering, Vol.**5**, Issue.**3**, pp.**92-97**, **June 2017.**

Authors Profile

Ms. A. D. Rewade pursed Bachelor of Engineering in Information Technology from SGB Amravati University of India in 2014 and pursuing Master of Technology in Computer Science and Engineering from RTM Nagpur University.

Dr. S. W. Mohod working as a Professor and Head of Department of Computer Engineering at Bapurao Deshmukh College of Engineering, Sevagram, Wardha, India. He has more than 25 years of teaching experience. He has published number of research papers in Reputed International Journals and conferences. His main research work focuses on Data Mining, Machine learning and Networking.

S. P. Bargat (M.Tech., MBA) working as a Assistant Professor in Department of Mechanical Engineering at Terna College of Engineering, Nerul, Navi Mumbai, India. He has more than 9 years of teaching experience. He has published number of research papers in Reputed International Journals and conferences. His main research area on Heat Transfer, Thermal Engineering and IC Engine.