IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

Content Based Alternate Medicine Recommendation By Using Random Forest Algorithm: A Review

Ankita D. Rewade[1*], Sudhir W. Mohod[2]

[1,2]Dept. of Computer Science Engineering, Bapurao Deshmukh College of Engineering, Nagpur University, Sevagram, Wardha, India

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

Available online at: www.ijcseonline.org

Accepted: 20/Oct/2018, Published: 31/Oct/2018

***Abstract* –** This paper gives the review of different prediction and recommendation system associated with health related problems. Increasing cost of medicine which is not affordable to generalized people and they are always looking for low cost medicine with same content and its effect is the main motivation behind this work. Alternate Medicine System solves this problem by searching through large volume of dynamically generated information to provide users with personalized content and services. This work gives the use of Random Forest Algorithm for content based alternate medicine recommendation system in order to serve as a useful tool for everyone who is associated with the medicine. This paper includes the different classification technique which recommends the different alternative solution. This work explores the different methods and technique used for prediction and recommendation of different issues regarding illness by using different classification techniques in recommendation systems. This study reveals the use of Random Forest Algorithm in the recommendation system. This is gives fast response and fast to build. It is even faster to predict and requiring cross-validation alone for model selection.

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

I. INTRODUCTION

Alternate Medicine System research has made significant advances over the past decades and has seen wide adoption in electronic commerce. Recently, a variety of types of side information (e.g., social friends, item content) has been incorporated into Alternate Medicine System to further enhance their performance, especially the well-recognized problem of data scarcity. However, most of existing approaches have only investigated the value of a single type of side information at a time, such as social trust, friendship, or item contents.

It is necessary to build new theories, techniques and methods to exploit multi-dimensional (homogeneous and heterogeneous) side information to provide users with better-personalized recommendations. At the same time, the large volume and variety of side data and the velocity of incremental updates in live systems provide challenges for the scalable mining and application of user preferences.
It is evident that the health of an individual significantly affects her quality of life. For this reason, finding appropriate physicians to diagnose and treat medical conditions is one of the most important decisions that a patient must make.

Currently, patients have two options that can aid them in addressing this problem, but both are of limited applicability. The first option is to rely on friends and family for advice on where to seek treatment. While recommendations produced by a close circle of friends can be assumed to be very trustworthy, the likelihood that friends and family have experience with the same medical history as the patient is quite low.

Furthermore, such advice can often be unavailable when, for instance, a patient moves to a new area and does not have an established network from which to seek advice; even when this is not the case, the number of physicians which friends and family have had contact with may not adequately cover the options in the given area. The second option for patients is to seek public information about and/or ratings for a physician available on, e.g., the internet. Such ratings, however, are sparse as medical history is often treated as personal, confidential information. Public ratings also suffer from the problem of trustworthiness, as the likelihood of inaccuracies is higher.

In order to combat the problem of a paucity of experience among a patient's trusted friends and the limited value of the

existing types of rating systems, propose a framework which enables patients to gather reliable doctor recommendations for their condition(s) while protecting the privacy of both (i) the patients contributing their ratings to the system and (ii) the patients making inquiries. In this framework patients can rate physicians based on their satisfaction (defined on a per condition basis) affording the patients more fine-grained control over how to choose the physician who best suits their needs.

It also protects the reliability of the results, meaning that (i) dishonest users can only minimally influence the outcome of a physician's rating and (ii) no physician (or small group of users) has the ability to tamper with the ratings. This enables the system to maintain the integrity of its ratings and ensure they are as unbiased as possible.

Motivation

The explosive growth in the amount of available digital information and the number of visitors to the Internet have created a potential challenge of information overload which hinders timely access to items of interest on the Internet. Information retrieval systems, such as Google, have partially solved this problem but privatization and personalization (where a system maps available content to user's interests and preferences) of information were absent. This has increased the demand for Alternate Medicine System more than ever before.

Alternate Medicine System are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to medicines based on their cost analysis and many different parameters.
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.

II. REVIEW OF LITERATURE

There has been considerable research into privacy preserving recommendation systems. Originally, privacy was achieved in recommendation systems by giving user information to a trusted third party, who then performs the necessary calculations with other trusted agents. One problem with this early approach is that, in addition to privacy, in order to be useful, recommendation systems must be robust against misbehaving users. One common way misbehaving users may attempt to influence the rating of a specific physician is known as "shilling attacks." Shilling attacks are said to occur when a user attempts to sabotage a competitor in order to make themselves look better.

Naoki Shino, Ryosuke Yamanishi, Junichi Fukumoto [1] has developed the recommendation system for alternative ingredients. The recommendation ingredients based on co-occurrence frequency of ingredients on recipe database and ingredient category stored in a cooking ontology. The result of the subjective evaluation experiments showed 88% appropriateness for alternative-ingredients recommendation.

Weiwei Lin, Ziming Wu, Longxin Lin, Angzhan Wen, And Jin Li. [2] has done the work on Ensemble Random Forest Algorithm for Insurance Big Data Analysis due to the imbalanced distribution of business data, missing user features, and many other reasons, directly using big data techniques on realistic business data tends to deviate from the business goals. It is difficult to model the insurance business data by classification algorithms, such as logistic regression and support vector machine (SVM). In this paper, author exploit a heuristic bootstrap sampling approach combined with the ensemble learning algorithm on the large-scale insurance business data mining, and propose an ensemble random forest algorithm that uses the parallel computing capability and memory-cache mechanism optimized by Spark. They collected the insurance business data from China Life Insurance Company to analyze the potential customers using the proposed algorithm. They use F-Measure and G-mean to evaluate the performance of the algorithm. Experiment result shows that the ensemble random forest algorithm outperformed SVM and other classification algorithms in both performance and accuracy within the imbalanced data, and it is useful for improving the accuracy of product marketing compared to the traditional artificial approach.

California's Workers' Compensation System (CAWCS) [3] has done the study of Department of Industrial Relations questioned the adequacy of the current Medi-Cal fee-schedule pricing and requested analysis of alternatives that maximize price availability and maintain budget neutrality. Authors has compare CAWCS pharmacy-dispensed (PD) drug prices under alternative fee schedules, and identify combinations of alternative benchmarks that have prices available for the largest percentage of PD drugs and that best reach budget neutrality. Author's claims transaction-level data from CAWCS were used to estimate total annual PD pharmaceutical payments. Medical pricing data was from the Workman's Compensation Insurance System (WCIS). Average Wholesale Prices (AWP), Wholesale Acquisition Costs (WAC), Direct Prices (DP), Federal Upper Limit (FUL) prices, and National Average Drug Acquisition Costs (NADAC) were from Medi-Span. Authors matched National Drug Codes (NDCs), pricing dates, and drug quantity for comparisons. Authors reported pharmacy-dispensed (PD) claims frequency, reimbursement matching rate, and paid costs by CAWCS as the reference price against all alternative price benchmarks.

Torgyn Shaikhinaa [4] has developed the Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. Clinical datasets are commonly limited in size, thus restraining applications of Machine Learning (ML) techniques for predictive modelling in clinical research and organ transplantation. They explored the potential of Decision Tree (DT) and Random Forest (RF) classification models, in the context of small dataset of 80 samples, for outcome prediction in high-risk kidney transplantation. The DT and RF models identified the key risk factors associated with acute rejection: the levels of the donor specific IgG anti-bodies, the levels of IgG4 subclass and the number of human leukocyte antigen mismatches between the donor and recipient. Furthermore, the DT model determined dangerous levels of donor-specific IgG subclass antibodies, thus demonstrating the potential of discovering new properties in the data when traditional statistical tools are unable to capture them. The DT and RF classifiers developed in this work predicted early transplant rejection with accuracy of 85%, thus offering an accurate decision support tool for doctors tasked with predicting outcomes of kidney transplantation in advance of the clinical intervention.

Shyong (Tony) K., Lam John Riedl [5] describe the attacks and discuss how they can affect the recommender system. Specifically, the authors consider various attack motivations (e.g., increasing/decreasing, the rating of an item and hindering the credibility of the recommendation system as a whole) and their effect on recommendation systems. Importantly, they note that while observing sharp changes in scores is an obvious way to detect (some) shilling attacks, non-trivial attacks against the system could potentially succeed. Detecting such attacks is proposed as a future area of research.

Bradley N. Miller [6] has worked toward a Personal Recommender System This position paper is an outgrowth from work reported in a Transaction on Information Systems article entitled Pocket Lens: Toward a Personal Recommender System, July 2004, in which we propose and compare several architectures for a decentralized recommender system built on top of peer-to-peer infrastructure. In this paper we review the need for personal recommender systems and propose the deployment of personal recommender systems using existing RSS and weblog technologies as the underlying communication infrastructure.

Deepti Sisodiaa, Dilip Singh Sisodiab [7] has worked on Diabetes as one of the deadliest and chronic diseases which causes an increase in blood sugar. Many complications occur if diabetes remains untreated and unidentified. The tedious identifying process results in visiting of a patient to a diagnostic center and consulting doctor. But the rise in machine learning approaches solves this critical problem. The motive of this study is to design a model which can prognosticate the likelihood of diabetes in patients with maximum accuracy. Therefore three machine learning classification algorithms namely Decision Tree, SVM and Naive Bayes are used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Indians Diabetes Database (PIDD) which is sourced from UCI machine learning repository. The performances of all the three algorithms are evaluated on various measures like Precision, Accuracy, F-Measure, and Recall. Accuracy is measured over correctly and incorrectly classified instances. Results obtained show Naive Bayes outperforms with the highest accuracy of 76.30% comparatively other algorithms. These results are verified using Receiver Operating Characteristic (ROC) curves in a proper and systematic manner.

Existing System

For product recommendation many systems are available in market for electronics good sale, grocery and clothing departments. But still now there are not systems available for providing alternates for medicines. In existing systems we have seen collaborative filtering and semantic signatures based recommended engines. Unfortunately this engine doesn't work efficiently for recommendation of medicines to users.

Limitation of Work

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.

III. PROPOSED SYSTEM

Front End:

[1] **Client Side:**
JSP-Web
Mobile UI based on bootstrap

[2] **Server Side:**
Apache Tomcat 7

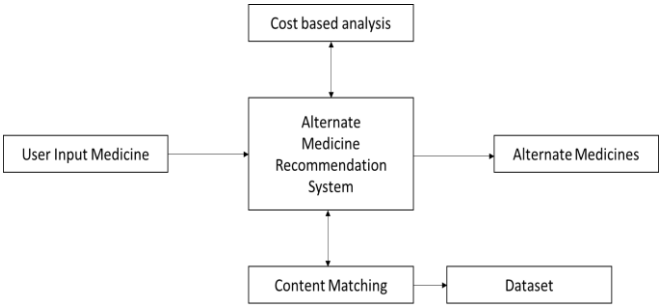
Back End:
MySQL 5.0

Software Requirement

Operating System: Window 8, Linux
Server : Apache Tomcat
Java : JDK 1.8
Database : MySQL 5.0
IDE : Eclipse
(Integrated Development Environment)

Hardware Requirement
Processor : i3 processor
RAM : 4 GB
Hard Disk : 256 GB

SYSTEM ARCHITECTURE



IV. ALGORITHM: RANDOM FOREST

Random Forest is an ensemble method in machine learning which involves construction (growing) of multiple decision tree (DT) via bootstrap aggregation (bagging). In other words, each time an input is supplied to RF that input is passed own each of the constituent decision tree. Each tree predicts a classification independently and "votes" for the corresponding class. The majority of the votes decide the overall RF prediction. This aggregate vote of several decision tree inherently less noisy and less susceptible to outliers than a single decision tree output, which mitigates the volatility due to small data and improves the robustness of predictions[5].

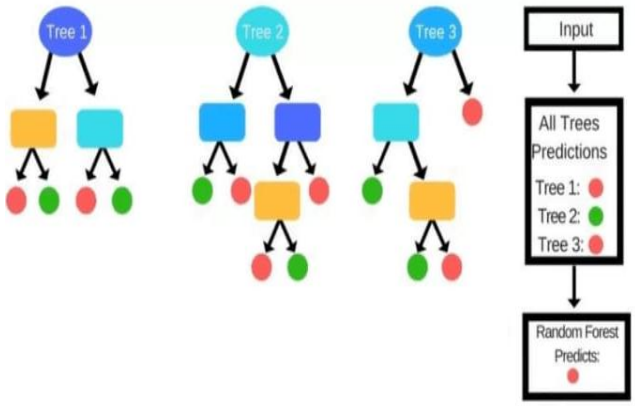


Figure 1. Decision Tree [5]

ADVANTAGES OF RANDOM FORESTS

- It can handle thousands of input variables without variable deletion
- It is one of the most accurate learning algorithms available
- For many data sets, it produces a highly accurate classifier
- It runs efficiently on large data bases
- Random forests are widely used

APPLICATIONS

- Health Care units at hospitals
- Pharmacy shops
- Individual Apps for public interest

DATA FLOW DIAGRAM

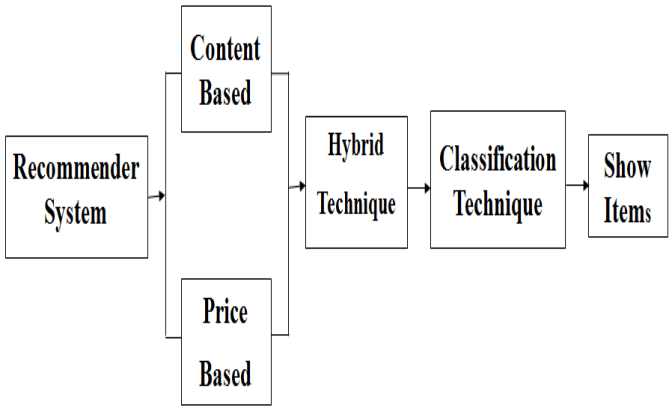


Figure 2. Data Flow Diagram

FLOW CHART

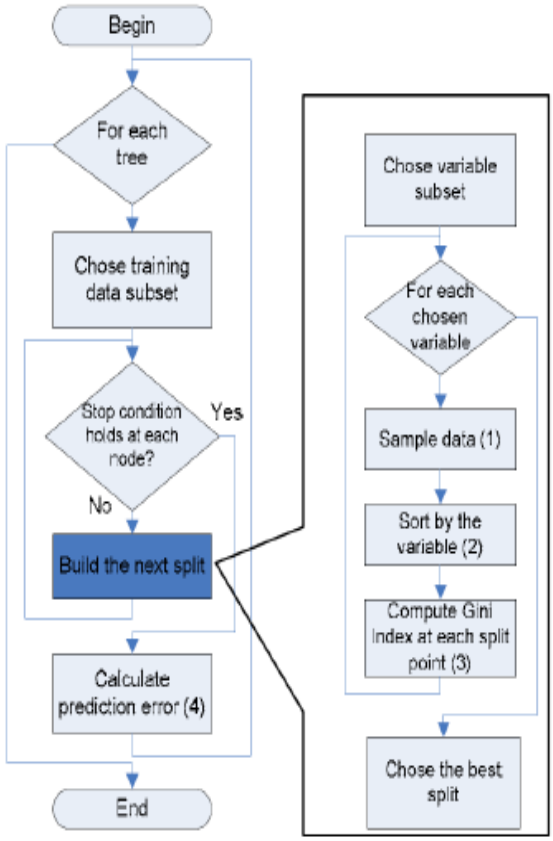


Figure 3. Flow Chart [5]

TECHNOLOGY REVIEW:

- **HTML**

Hypertext Mark-up Language (HTML) is the standard mark-up language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript, it forms a triad of cornerstone technologies for the World Wide Web.[4] Web browsers receive HTML documents from a web server or from local storage and render them into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects, such as interactive forms, may be embedded into the rendered page. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as <img /> and <input /> introduce content into the page directly. Others such as ... surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.

HTML can embed programs written in a scripting language such as JavaScript which affect the behavior and content of web pages. Inclusion of CSS defines the look and layout of content. The World Wide Web Consortium (W3C), maintainer of both the HTML and the CSS standards, has encouraged the use of CSS over explicit presentational HTML since 1997.

- **Cascading Style Sheets**

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language.[1] Although most often used to set the visual style of web pages and user interfaces written in HTML and XHTML, the language can be applied to any XML document, including plain XML, SVG and XUL, and is applicable to rendering in speech, or on other media. Along with HTML and JavaScript, CSS is a cornerstone technology used by most websites to create visually engaging web pages, user interfaces for web applications, and user interfaces for many mobile applications.

CSS is designed primarily to enable the separation of presentation and content, including aspects such as the layout, colors, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple HTML pages to share formatting by specifying the relevant CSS in a separate .css file, and reduce complexity and repetition in the structural content.

Separation of formatting and content makes it possible to present the same mark-up page in different styles for different rendering methods, such as on-screen, in print, by voice (via speech-based browser or screen reader), and on Braille-based tactile devices. It can also display the web page differently depending on the screen size or viewing device. Readers can also specify a different style sheet, such as a CSS file stored on their own computer, to override the one the author specified.

- **MySQL**

MySQL is an open-source relational database management system (RDBMS).[7] Its name is a combination of "My", the name of co-founder Michael Widenius's daughter and "SQL", the abbreviation for Structured Query Language. The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation. For proprietary use, several paid editions are available, and offer additional functionality. MySQL is a central component of the LAMP open-source web application software stack (and other "AMP" stacks). LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python". Applications that use the MySQL database include: TYPO3, MODx, Joomla, WordPress, Simple Machines Forum, phpBB, MyBB, and Drupal. MySQL is also used in many high-profile, large-scale websites.

V. CONCLUSION

Medicine availability and costing are two major issues in health care system these days. Many times the medicines prescribed by doctor are not available or are very costly based on brand names. Recently many new medicine developers have stepped into market with low cost medicines with high availability. Our proposed system is going to provide users with such medicine information that are less costly as compared to prescribed medicines and at the same time have high availability in market thus helping users in avoid higher cost in health care. 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 and Automatic handling of missing values. It gives the best alternate predictor selection from large number of medicine. In the future alternate medicine recommendation system will be develop by using random forest algorithm. The limitation of the system is only the available medicine in the prepared data set will recommended to the 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, DES's College of Engineering & Technology, Dhamangaon Rly., Amravati 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.

