

## A Survey on HealthBot using Machine Learning Algorithms

N. Sathya<sup>1\*</sup>, S. Jayendren<sup>2</sup>, S. Prasanth<sup>3</sup>, S. Sharmila<sup>4</sup>

<sup>1,2,3,4</sup>Department of Information Technology, Sri Shakthi Institute of Engineering and Technology, Anna University, Coimbatore, India

\*Corresponding Author: [nsathyait@siet.ac.in](mailto:nsathyait@siet.ac.in) Tel.: 9698460166

DOI: <https://doi.org/10.26438/ijcse/v7i2.743749> | Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 15/Feb/2019, Published: 28/Feb/2019

**Abstract**— Nowadays, People care a lot about their health. Similarly, they have a pocket rocket technical object, which depicts the technological updating of the well beings. There is an application called HealthBot. HealthBot is a computer program planned to re-enact discussion with human clients to procure information approximately his wellbeing status, especially over the Internet. The Wellbeing Bot Benefit could be a SaaS arrangement that engages its accomplices to construct and send compliant, AI-powered wellbeing operators, permitting them to offer their clients brilliantly, personalized get to health-related data and intuitive through a normal discussion encounter. So, on considering the people's priority list, an application that takes care of one's health with the help of that technology was considered into account. As an account of such consideration, Rovi 1.0 is born from the three Budding Software engineers. At present, we have such things but this rovi will stand out in many aspects. The main scope & objective of this is to act like a packet nurse. To suggest the medicine for the diseases. It includes the most efficient way of maintaining the report that is useful in any situation. It includes the main supervised learning process that is more efficient. Prevention is better than cure. Systems are needed for dynamic interaction with people to gather Information, Monitor health condition and provide support, especially after hospital discharge or at-home settings. This can be possible by using the healthbot technique. This paper highlights different Information Mining strategies such as classification, clustering, affiliation conjointly highlights related work to dissect and foresee human illness.

**Keywords**— Random forest algorithm, classification algorithm, clustering algorithm, K-mean, support vector machine.

### I. INTRODUCTION

Nowadays is the time of insights in machines. With the progresses in Manufactured Insights, machines have begun to mimic distinctive human characteristics nowadays. A chatterbot or chatbot points to create a discussion between both human and machine. By comparison, machine learning arrangements approach issues in a way more closely taking after a recently authorized therapeutic inhabitant as she learns unused data and rules from information input over time. With "experience," machine learning calculations can seek for combinations and designs and dependably anticipate results by dissecting expansive sums of information. This prepare is comparative to conventional relapse models – where factors that matter are sorted from those that can be disregarded. The machine has been inserted information to recognize the sentences and making a choice itself as reaction to reply an address. The reaction guideline is coordinating the input sentence from client. The information of chatbot are put away within the database. The chatbot comprises of center and interface that's getting to that center in social database administration frameworks (RDBMS). The database has been utilized as information capacity and

translator has been utilized as put away programs of work and procedure sets for pattern-matching prerequisite. Machine plays a major part within the advancement of innovation. The productive utilization of machine is to interaction with the machine by inquiring different inquiries to computer and legitimate reaction from the machine.

Information building could be a center portion of AI inquiries about. Machines can regularly act and respond like people as it were on the off chance that they have plenteous data relating to the world. Manufactured insights must have got to objects, categories, properties and relations between all of them to execute information building. Starting common sense, reasoning and problem-solving control in machines may be a troublesome and dull assignment.

Machine learning is additionally a center portion of AI. Learning without any kind of supervision requires a capacity to distinguish designs in streams of inputs, while learning with satisfactory supervision includes classification and numerical relapses. Classification decides the category a protest has a place to and relapse bargains with getting a set

of numerical input or yield illustrations, in this manner finding capacities empowering the era of appropriate yields from individual inputs. Scientific investigation of machine learning algorithms and their execution may be a well-defined department of hypothetical computer science regularly alluded to as computational learning theory. Machine discernment bargains with the capability to utilize tangible inputs to infer the distinctive aspects of the world, whereas computer vision is the control to analyse visual inputs with a number of sub-problems such as facial, question and motion recognition. Robotics is additionally a major field related to AI. Robots require insights to handle all the major parts.

## II. LITRATURE SURVEY

[1] Wellbeing care industry produces colossal amount of information that clutches complex data relating to patients and their restorative conditions. Information mining is picking up notoriety completely different inquire about fields due to its unbounded applications and strategies to mine the data in redress way. Information mining techniques have the capabilities to find covered up designs or connections among the objects in the medical information. In final decade, there has been increment in utilization of information mining methods on therapeutic data for deciding valuable patterns or designs that are utilized in investigation and choice making. Information mining has an unbounded potential to utilize healthcare information more effectively and solidly to anticipate distinctive kind of disease. This paper highlights different Information Mining strategies such as classification, clustering, association and moreover highlights related work to dissect and foresee human disease.

The pre-proposed systems have the ability to suggest the medicines for the general diseases which are frequent to the area where the bot was created. They tend to exist like the HIKE Natasha. Some of the examples of the successful healthbots in the society are Babylon, Izzy, etc.

Information mining is a collection of algorithmic procedures to extricate teacher designs from crude information. Healthcare industry nowadays produces gigantic sums of diverse information around healing centers, assets, infection determination, electronic persistent records, etc. The huge sum of information is pivotal to be handled and scrutinized for information extraction that enables back for understanding the winning circumstances in healthcare industry. Information mining forms incorporate surrounding a speculation, gathering information, performing pre-processing, evaluating the demonstrate, and understanding the show and draw the conclusions. Information mining procedures such as affiliation, classification and clustering are utilized by healthcare association to extend their capability for building the fitting conclusion with respect to the patient's wellbeing.

In common, Information Disclosure (KDD) and Data Mining are related terms and are utilized traded but a few analysts expect that both terms are different as Information Mining is one of the foremost imperative stages of the KDD process. According to Fayyad et al., the Information Disclosure in database is systematized in various stages while the primary arrange is choice of information in which information is accumulated from distinctive sources, the moment organize is pre-processing the chosen information, the third organize is changing the information into suitable organize so that it can be prepared assist, the fourth arrange comprise of Information Mining where suitable Information Mining procedure is connected on the changed information for extricating valuable information and assessment is the final stage The primary drawback of the existing framework is that these Frameworks does not have a report producing capability such that the individuals may feel unsatisfied with the unstable get to their past inputs. They have no capacity to compare the cure.

[2] Random forest is an outfit of classification calculation broadly utilized in much application particularly with bigger datasets since of its exceptional highlights like Variable Significance degree, OOB mistake location, Nearness among the include and dealing with of imbalanced datasets. This paper examines numerous applications which utilize Random forest to classify the dataset like Organize interruption discovery, Mail spam location, quality classification, Credit card extortion discovery, and Content classification. In this paper each application is briefly presented and after that the dataset utilized for usage is examined and at last the genuine execution of this algorithm with steps shrewd strategy additionally the comes about are examined. This algorithm Calculation and its highlights are too talked about to highlight many Calculation more clearly.

Features of Random Forests:

1. It is unexcelled in accuracy among current algorithms.
2. It runs efficiently on large data bases.
3. It can handle thousands of input variables without variable deletion.
4. It gives estimates of what variables are important in the classification.
5. It generates an internal unbiased estimate of the generalization error as the forest building progresses.
6. It has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing.
7. It has methods for balancing error in class population unbalanced data sets.
8. Generated forests can be saved for future use on other data.

9. Prototypes are computed that give information about the relation between the variables and the classification.
10. It computes proximities between pairs of cases that can be used in clustering, locating outliers or (by scaling) give interesting views of the data.
11. The capabilities of the above can be extended to unlabeled data, leading to unsupervised clustering, data views and outlier detection.
12. It offers an experimental method for detecting variable interactions.

Random forest is a directed machine learning calculation. In Information Mining space, machine learning calculations are broadly utilized to analyze information, and create forecasts based on this information. Being an outfit calculation, Random Forest creates different choice as base classifiers and applies larger part voting to combine the results of the base trees. Quality of person choice trees and relationship among the base trees are key issues which choose generalization mistake of Random Forest classifiers. Based on precision degree, Random Forest classifiers are at standard with existing outfit strategies like sacking and boosting. In this investigate work an endeavor is made to progress execution of Random Forest classifiers in terms of precision, and time required for learning and classification. To attain this, five modern approaches are proposed. The experimental examination and results of tests carried out in this investigate work lead to successful learning and classification utilizing Random forest algorithm.

[3] A number of directed learning calculations are by and by being utilized for different applications. Most works will either center on the execution of a certain calculation or analyse distinctive classification procedures. One of the numerous centers of the paper is the right choice of classifiers and parameters in practical circumstances has been a long-standing issue. The point is to compare and recognize calculations over different dimension including computational execution. The paper evaluates the execution of Back Vector Machines, Naïve Bayes and Choice trees. Machine learning for the most part known as ML is variety of machine Insights (AI) which comprise of computers with the capability to be prepared while not being explicitly modified. Machine learning centres consideration on the development of pc programs that has sufficient capability to alter once unprotected to unique information. ML calculations are freely classified into 3 divisions unsupervised learning, administered learning and reinforcement learning. Machine learning like information mining has advanced significantly in later a long time. Machine learning like Information mining aim at analysing the total information and attempt to discover sensible designs in it. A number of directed learning calculations are directly being utilized for different applications. Most works will either centre on the execution

of a certain calculation or analyse different classification procedures. One of the numerous centres of the paper is the correct determination of classifiers and parameters in down to earth circumstances have been a long-standing issue. The point is to compare and recognize calculations over different measurement counting computational execution. The paper evaluates the execution of Support Vector Machines.

The Machine Learning (ML) field has picked up its force in nearly any domain of inquire about and just as of late has gotten to be a dependable tool within the medical domain. The observational domain of programmed learning is utilized in assignments such as therapeutic choice support, therapeutic imaging, protein-protein interaction, extraction of therapeutic information, and for in general persistent administration care. ML is imagined as a instrument by which computer-based systems can be coordinates within the healthcare field in arrange to get an improved, more productive therapeutic care. This paper describes a ML-based strategy for building an application that's competent of distinguishing and spreading healthcare data. It extracts sentences from published therapeutic papers that specify diseases and treatments, and recognizes semantic relations that exist between diseases and treatments. Our evaluation results for these assignments appear that the proposed methodology gets dependable results that might be coordinates in an application to be utilized within the therapeutic care domain. The potential esteem of this paper stands within the ML settings that we propose and within the truth that we outperform past comes about on the same information set.

**Data Heterogeneity:** Once the features vectors consist of features of many sorts that has separate, separate ordering, frequency, continuous values, and bound algorithmic expression are easier to use than remainder rest algorithms. Several different of these algorithms particularly - regression, SVM, supply regression, nearest neighbour ways and ANN needs the input features to not be nominal should exist in numerical form and scaled accordingly.

**Data Redundancy:** Once the data that is provided options has undesirable data, some learning rule in all probability could execute defectively attributable to numerical irresolution. Such researches problems are also resolved by processing the data before using it.

**Presence of interactions and non-linearity's:** Once the feature has an independent part to the output, then the rue that supported distance functions and linear functions typically achieve match. Whereas the opposite hand, once we have multifarious communications between feature, and bound rule achieve far enhanced results, as they are typically intended to work out these communications.

### III. PROPOSED SYSTEM ALGORITHMS

Assist the common people in maintaining their health. It also maintained the Medical Report for individual Provide prescription for general and frequent diseases. ROVI denotes the name of the healthbot. The Day-to-Day feed about one's body condition to rovi helps you to generate a report of your health. Rovi will also alarms you when you are about to have a serious deadly disease based on the inputs provided to it. Rovi reads the provided inputs and store them for future purpose. Rovi is said to be built with the capability to alert you about the serious disease ahead in your near coming future.

The main scope & objective of this is to act like a packet nurse. To suggest the medicine for the diseases. It includes the most efficient way of maintaining the report that is useful in any situation. It includes the main supervised learning process that is more efficient. In the current system there are no such methods like reporting and the self-learning by the current input data from the use is not available. This system overcomes all those disadvantages.

#### Knowledge Discovery:

Knowledge Discovery in Databases, or KDD for brief, refers to the wide handle of finding knowledge in information, and emphasizes the "high-level" application of specific information mining strategies. It is of interest to analyst's machine learning, design acknowledgment, databases, measurements, artificial intelligence, information acquisition for master frameworks, and information visualization. The binding together objective of the KDD handle is to extract information from information within the setting of expansive databases. It does this by utilizing information mining strategies (calculations) to extract (distinguish) what is regarded information, according to the determinations of measures and thresholds, employing a database at the side any required pre-processing, subsampling, and changes of that database.

**KDD** refers to the overall process of finding valuable information from information. It includes the assessment and possibly interpretation of the designs to create the choice of what qualifies as information. It also incorporates the choice of encoding plans, pre-processing, testing, and projections of the information earlier to the information mining step.

**Data mining** refers to the application of calculations for extracting designs from information without the extra steps of the KDD process.

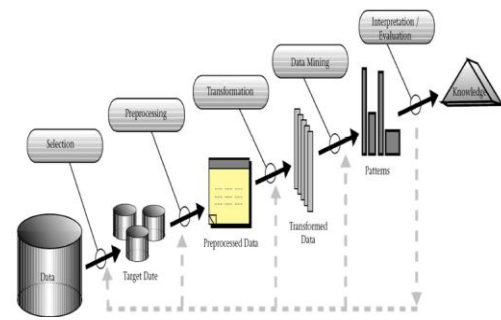


Figure 1. Stages of Knowledge Discovery Process

The overall process includes,

1. Developing an understanding of
    - i) The application domain
    - ii) The significant earlier knowledge
    - iii) The objectives of the end-user
  2. Creating a target data set: selecting a data set, or focusing on a subset of variables, or data samples, on which discovery is to be performed.
  3. Data cleaning and pre-processing.
    - i) Removal of noise or outliers
    - ii) Collecting fundamental data to show or account for noise
    - iii) Strategies for dealing with lost information fields
    - iv) Accounting for time grouping data and known changes
  4. Data reduction and projection.
    - i) Finding valuable highlights to speak to the Information depending on the objective of the task
    - ii) Using dimensionality decrease or change strategies to decrease the viable number of factors under thought or to discover invariant representations for the data
  5. Choosing the information mining tasks.
    - i) Deciding whether the objective of the KDD prepare is classification, regression, clustering, etc.
  6. Choosing the information mining algorithms.
    - i) Selecting method(s) to be utilized for looking for designs within the data.
    - ii) Choosing which models and parameters may be suitable
    - iii) Matching a specific information mining strategy with the by and large criteria of the KDD process.
  7. Data mining.
    - i) Searching for designs of interest in a specific representational frame or a set of such representations as classification rules or trees, regression, clustering, and so forth.
  8. Interpreting mined patterns.
  9. Consolidating found knowledge
- Algorithm/ Technology Description:

#### i) Random forest algorithm:

Random Forest is an administered machine learning calculation. In Data Mining domain, machine learning

calculations are broadly utilized to analyse information, and produce forecasts based on this information. Being a gathering calculation, Random Forest creates different choice trees as base classifiers and applies majority voting to combine the results of the base trees. Quality of person choice trees and relationship among the base trees are key issues which choose generalization mistake of Random Forest classifiers. Based on exactness degree, Random Forest classifiers are at standard with existing gathering procedures like stowing and boosting. In this investigate work an endeavour is made to improve execution of Random Forest classifiers in terms of exactness, and time required for learning and classification. To realize this, five modern approaches are proposed. The observational examination and results of experiments carried out in this investigate work lead to compelling learning and classification using the random forest algorithms.

Random Forest (RF) is an outfit, supervised machine learning algorithm. Machine learning procedures are connected within the domain of Information Mining. A Random Forest could be a classifier comprising of a collection of tree-structured classifiers  $\{h(x, \Theta_k) \mid k=1, 2, \dots\}$ , where the  $\{\Theta_k\}$  are free unclearly disseminated random vectors and each tree casts a unit vote for the most well-known lesson at input  $x$  [6]. Random Forest produces an outfit of choice trees. To produce each single tree in Random Forest, Breiman taken after taking after steps: In case the number of records within the training set is  $N$ , at that point  $N$  records are examined at irregular but with substitution, from the first information; usually bootstrap test. This test will be the preparing set for developing the tree.

Random Forest implies the forest of choice trees produced utilizing Forest RI prepare. Within the woodland building handle, when bootstrap test set is drawn by examining with substitution for each tree, approximately 1/3rd of unique occurrences are cleared out. This set of occurrences is called OOB (Out-of-bag) information. Each tree has its claim OOB information set which is utilized for error estimation of person tree within the forest, called as OOB mistake estimation. The Generalization mistake of Random Forest is given as,  $PE^* = P_{x,y} (mg(X,Y)) < \text{The edge work is given as, } mg(X,Y) = \text{avk } I(hk(X) = Y) - \max_{j \neq Y} \text{avk } I(hk(X) = j)$ . The edge work measures the degree to which the normal number of votes at  $(X, Y)$  for the proper course exceeds the normal vote for any other course. Quality of Random Forest is given in terms of the anticipated esteem of edge work as,  $S = E_{X, Y} (mg(X, Y))$ . In case  $\rho$  is mean esteem of relationship between base trees, an upper bound for generalization error is given by,  $PE^* \leq \rho (1 - s^2) / s^2$ . Subsequently, to surrender superior precision in Random Forest, the base choice trees are to be assorted and exact.

#### ii) *K-Mean algorithm:*

K-implies grouping is a technique for vector quantization, initially from flag preparing that is well known for bunch investigation in information mining. k-implies grouping plans to parcel  $n$  perceptions into  $k$  bunches in which every perception has a place with the bunch with the closest mean, filling in as a model of the group. This outcome in a dividing of the information space into Voronoi cells. The issue is computationally troublesome (NP-hard); in any case, there are effective heuristic calculations that are normally utilized and meet rapidly to a neighbourhood ideal. These are normally like the desire boost calculation for blends of Gaussian disseminations through an iterative refinement approach utilized by both k-implies and Gaussian blend displaying. Furthermore, they both utilize bunch focuses to demonstrate the information; nonetheless, k-implies grouping tends to discover groups of amount spatial degree, while the desire boost instrument enables bunches to have distinctive shapes. The calculation has a free relationship to the k-closest neighbour classifier, a well-known machine learning procedure for order that is regularly mistaken for k-implies because of the  $k$  in the name. One can apply the 1-closest neighbour classifier on the bunch focuses got by  $k$ -intends to arrange new information into the current groups.

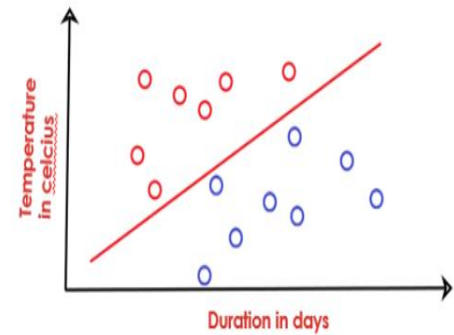
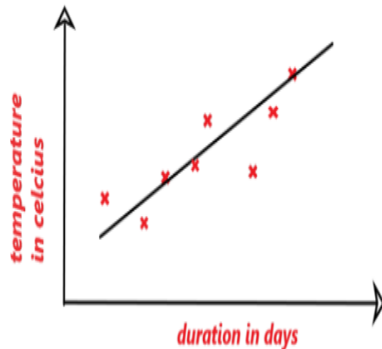
Given an arrangement of perceptions  $(x_1, x_2, \dots, x_n)$ , where every perception is a  $d$ -dimensional genuine vector, k-implies bunching means to parcel the  $n$  perceptions into  $k$  ( $\leq n$ ) sets  $S = \{S_1, S_2, \dots, S_k\}$  in order to limit the inside group entirety of squares (WCSS) (i.e. fluctuation). Formally, the goal is to find:

(1) Where  $\mu_i$  is the mean of focuses in  $S_i$ . This is proportionate to limiting the pairwise squared deviations of focuses in a similar group

(2) The aggregate fluctuation is consistent, this is likewise proportionate to boosting the squared deviations between focuses in various groups.

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mu_i\|^2 = \arg \min_{\mathbf{S}} \sum_{i=1}^k |S_i| \text{Var } S_i$$

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \frac{1}{2|S_i|} \sum_{\mathbf{x}, \mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2$$

**Advantage:**

1) If factors are gigantic, at that point K-Means the greater part of the circumstances computationally quicker than various levelled grouping, on the off chance that we keep k smalls.

2) K-Means deliver more tightly bunches than various levelled grouping, particularly if the groups are globular.

**Disadvantage:**

1) Difficult to foresee K-Value.

2) With worldwide bunch, it didn't function admirably.

3) Different introductory segments can bring about various last groups.

4) It doesn't function admirably with bunches (in the first information) of Different size and Different thickness

**iii) Support Vector Machine:**

In machine learning, bolster vector machines (SVMs, likewise bolster vector networks) are administered learning models with related learning calculations that break down information utilized for grouping and relapse examination. Given a course of action of planning outlines, each set separated as having a put with either of two classifications, a SVM planning calculation makes a show that designates unused cases to one class or the other, making it a non-probabilistic parallel direct classifier (in spite of the fact that strategies, for example, Platt scaling exist to utilize SVM in a probabilistic characterization setting). An SVM demonstrate is a portrayal of the cases as focuses in space, mapped with the goal that the cases of the different classifications are partitioned by a reasonable hole that is as wide as would be prudent. New illustrations are then mapped into that same space and anticipated to have a place with a class in light of which side of the hole they fall.

**Advantages:**

1. Firstly it has a regularisation parameter, which makes the user think about avoiding over-fitting.

2. Secondly it uses the kernel trick, so you can build in expert knowledge about the problem via engineering the kernel.

3. Thirdly a SVM is characterized by a raised enhancement issue (no nearby minima) for which there are productive strategies (e.g. SMO).

4. Lastly, it is an approximation to a bound on the test error rate, and there is a substantial body of theory behind it which suggests it should be a good idea.

**Disadvantages:**

1. Maybe the greatest impediment of the help vector approach lies in decision of the bit.

2. A second restriction is speed and size, both in preparing and testing.

3. Although SVMs have great speculation execution, they can be horrifyingly moderate in test stage.

**iv) Background Methods:**

As the spiking development of therapeutic information due to the rise of wellbeing data innovation within the past decades, it would be a tragic lost on the off chance that we don't create a compelling and application calculation to utilize the sea of therapeutic information. Choice tree such as ID3 is one of the foremost utilized classification strategies. C4.5 and C5.0 are improved adaptations of ID3. Random forest is an extended choice tree method. It was made to resolve the mistake proliferation issue related with conventional choice. This study employed the algorithm of C5.0 decision tree and random forest to analyse and predict health care problems.

Healthcare database was downloaded from the centre for machine learning and intelligent system, University of California, Irvine. Original dataset has 16 different classifications. In the study, we combined all the cases and classify the data into only two categories: normal (245 cases) and abnormal (the remaining 207 cases). Both the C5.0 decision tree and the random forest were done in Python platform. Sometime recently the preparing of information, the database was rearranged some time recently picking the

primary 350 information focuses as the preparing dataset. Completely, three choice trees based on C5.0 and three random forests were built.

#### IV. CONCLUSION AND FUTURE STUDY

Health is wealth. Need to protect our environment in order to protect our health. Be that as it may, the affectability and specificity are still not tall sufficient compare to doctor's conclusion, indeed in spite of the fact that we only classify them in two categories: typical and irregular. In arrange to apply to clinics, the calculations ought to be moved forward. Possible heading may be to consult to experienced specialists to combine expert system with machine learning. It can offer assistance to preselect qualities some time recently the information training. Another course to improve random forest is to look at each tree within the random forest to look at the heterogeneity of each tree, at that point giving weight within the last choice phase.

The subject has created considerable advancement within the past few years. Method of learning like random forests, boosting, bagging, and Support Vector Machines accomplished wonderful results that may are tough to get simply fifteen years earlier. One in the sooner learning ways, fed forward neural networks had the most effective results and square measure that is comparative with a number of the methods that are new, notably if algorithms won't mark once coaching.

#### REFERENCES

- [1] Sheetal L. Patil, "Survey Of Data Mining Techniques In Healthcare", International Research Journal of Innovative Engineering, Volume: 01 Issue: 09, September-2015.
- [2] "Classification of large datasets using Random Forest Algorithm in various applications: Survey" .Mohammed Zakariah Researcher, King Saud University College of Computer and Information Sciences, Riyadh, Kingdom of Saudi Arabia
- [3] International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, May 2014 "Effective Learning and Classification using Random Forest Algorithm", 1Vrushali Y Kulkarni, 2 Pradeep K Sinha
- [4] "Survey on Comparison of Supervised Learning Classification Algorithms In-depth Study of various Factors that Affect the Speed, Accuracy and Precision of Supervised Learning Algorithms", 1Umair Ahmed and 2T. Manoranjitham.
- [5] "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modelling". BayuSetiaji; Ferry Wahyu Wibowo.
- [6] "Paging Dr. Bot" – The Emergence of AI and Machine Learning in Healthcare. Michael Woolf, Baker, Donelson, Bearman, Caldwell & Berkowitz, PC, Nashville, TN.
- [7] C. A. Caligtan, P. C. Dykes, "Electronic health records and personal health records", *Semin Oncol Nurs*, vol. 27, pp. 218-228, 2011.
- [8] C.M. Bishop, "Pattern recognition and machine learning", New York: Springer, 2006.
- [9] "A survey of machine learning algorithms for big data analytics". S. Athmaja; M. Hanumanthappa; Vasantha Kavitha
- [10] International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016doi: 10.5121/Ijist.2016.6206 53. "Survey Of Data Mining Techniques Used In Healthcare Domain Sheenal Patel and Hardik Patel"
- [11] Basic And Weighted K-NN, Innovations in Intelligent Systems and Applications (INISTA),2012. Dhanya P Varghese &Tintu P B, "A Survey On Health Data Using Data Mining Techniques", International Research Journal of Engineering and Technology (IRJET), Volume:02 Issue: 07, Oct-2015.
- [12] "Ensemble Methods for Classification Of Patients For Personalized Medicine With High-Dimensional Data". Artificial Intelligence in Medicine 41:197–207,2007.

#### Authors Profile

##### N.SATHYA

Assistant Professor,  
Department of Information Technology ,  
Sri Shakthi Institute of Engineering and Technology,  
Coimbatore Tamilnadu.  
nsathyait@siet.ac.in



##### S.JAYENDREN,

B.Tech, Department of Information Technology ,  
Sri Shakthi Institute of Engineering and Technology,  
Coimbatore Tamilnadu.  
Jaijaish98@gmail.com



##### S.PRASANTH,

B.Tech, Department of Information Technology ,  
Sri Shakthi Institute of Engineering and Technology,  
Coimbatore Tamilnadu.  
prasantstaz@gmail.com



##### S.SHARMILA,

B.Tech, Department of Information Technology ,  
Sri Shakthi Institute of Engineering and Technology,  
Coimbatore Tamilnadu.  
sharmila.s2797@gmail.com

