xIlkogretim Online - Elementary Education Online,2021; Vol 20 (Issue 1): pp.2372-2382
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.01.265

A New And Fast Supervised Learning Algorithm Based On Blood Pressure (Bp) Data Analysis

S.Sivakumar, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation (KLEF), Guntur District, Andhra Pradesh, India - 522 502, Sivaiit79@gmail.com
S. Vidyanandini, Department of Mathematics, SRM Institute of Science and Technology, Kattankulathur , Chennai - 603 203.
E. Rajesh Kumar, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation (KLEF), Guntur District, Andhra Pradesh, India - 522 502.
D. Haritha, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation (KLEF), Guntur Dt, Andhra Pradesh, India - 522 502.
CMAK ZeelanBasha, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation (KLEF), Guntur District, Andhra Pradesh, India - 522 502.

Abstract- The supervised learning algorithm is one of the most popular techniques in data mining, Billions of business users and industries can use the fast classifier mining algorithm for classifying the data. This algorithm is tested in medical data sets for blood pressure (BP) which uses generic sorting techniques (quick sort) in the tree- growing segment. The classifier is suitable for handling both categorical and numerical attributes. The implementation and experimental study denote to classification problem where the main aim is to expect the split to classify them into low BP, high BP, and normal BP opinion with the aim of identifying attributes. The classification method excels in the case of handling a large set of data and attributes for medical data sets, which have been researched in current years with varying results. Presently object-oriented design of a fast classifier mining algorithm has been executed in java programming. This paper's motivation on an effective quick sort algorithm executed in java for the decision tree classifier is known as a fast classifier mining algorithm. This algorithm accumulates less process time and produces accurate and fast outcomes from the allocation of the node count values.

Keywords: blood pressure (BP), fast classifier mining algorithm, numerical attributes

I. INTRODUCTION

Data mining usually includes techniques such as Artificial Neural Networks [1]. Decision tree classifiers, Genetic algorithm, Association rule, and Clustering [5]. An essential tool that is regularly utilized for classification methods in data mining [2]. Various ways to deal with construct accurate classifiers have been proposed (e.g. ., Bayesian classifier [8], Decision tree [9], Support Vector Machine (SVMs) [10], and associative classifiers [11]). In classification model training dataset are used with records containing different attributes. Attributes like categorical and numerical attributes can be used in training dataset, if attribute values are in proper order according to the domain space is said to be numerical attributes (age, sports, sleep, drink, and weight). If the attribute values are in unordered manner, then it's known as categorical attribute (BP, sex) [5]. Classification is a method of separating a dataset into general limited group's names as class based on some attributes [4].

These techniques are primarily divided into 2 separate groups. 1. Decision tree (DT) constructed algorithm and 2.rule-based algorithm. The DT is one of the machine learning and data mining methods with a scope of uses in several fields [8]. DT is a tree structure demonstration of the specified DT, so that each internal node is connected with single decision variables to some extent, each branch from an internal node is related with corresponding decision variable value subset and every leaf node is related with an estimation of the objective variable [17]. The decision tree is developed from top to the bottom of the tree or root node to the terminal node. Rule-based algorithms can create various multiple splitting points of the idea [13]. In this algorithm the found information is frequently communicated as:

IF < conditions > THEN < (class value) >

We can apply rule based algorithm to each nodes and if the condition is false, then move to right side of if the node or the condition said as true then, move or go to the left side of the node else count the missing values[19]. This rule will travel up to *n* attributes, to end counts the class value. There are several display procedures for data in classification algorithm that are utilized as outcome formats [15]. Decision rules follows the simplest output formats: 'If Condition Then class. Here the ''IF" part is composed of attribute-value pairs with AND logical operators' used for splitting the multiple attributes and the "Then" part reveals the related counts of the class value in the specified form [16].

This paper motivation on the productivity quick sort algorithm when executed in java for the decision tree classifier is known as fast classifier mining algorithm. This method produces accurate outcomes is fast with less handling time and allocation of the node count value is constructed on the rule based algorithm. To end with construction of the decision tree and study of experimental outcomes from node distribution count values are obtained [14].

This paper is separated into different sections which are as per the following. Section 2 appropriately expresses the related works addressed in this paper. Section 3 offers the proposed method to deal with object-oriented plan for a fast classifier mining technique and implementation [7]. Section 4 methodically defines the case study for the medical database in BP and well-known histograms in all the nodes and travelled path. In section 5 a new evaluation outcome value of the proposed technique is explained. Section 6 conclusions and discusses future works [6].

II. RELATED WORKS

In this study mostly constructed on three methods, Supervised Learning in Quest (SLIQ) is based on classifier DT that can deal with both numerical attributes and categorical attributes. It constructs accurate and reduced trees. It utilizes a new pre-sorting method in the tree growing stage to minimize the rate of calculating numerical attributes [3]. This methodology is incorporated with a breadth-first tree (BFT) developing approach to empower classification of datasets that are disk-resident. To determine the split of categorical attribute SLIQ uses an algorithm: fast-sub-setting. Scalable Parallelizable Induction of decision Trees (SPRINT)prearranged classification algorithm is also known as SPRINT that takes out all constraints related to memory that restrict DT method that demonstrates that technique once planned are scalable and fast [6].

A new proposed method for dealing along with categorical attributes and numerical attributes in larger datasets called as Mixed Mode Database Miner (MMDBM) issued. The proposed model deal through huge database with a large arrangement of data or attributes. A new well-organized index to qualify splitting points is introduced [12]. This are separated into two parts, initial one is predictive classifier gives a complete explanation of our method and another model is object oriented design (OOD), provides the object-oriented(OO)execution of proposed procedure and explanation of the front-end established in this model. In this section, the researcher has compared their technique with recognized SPRINT, MMDBM and SLIQ algorithm. In proposed model it is recommended that DT classifiers like SLIQ, SPRINT and MMDBM have accomplished a better accuracy [18].

A. Decision tree Algorithm

DT algorithm is a supervised learning technique utilized in data mining for regression and classification techniques. Decision tree is based on a tree concept, which helps us in decision-making purposes. A DT is a model that combines branch, root node, and leaf or terminal nodes. Every leaf node indicates a test on every single attribute, every branch indicates the test result, and each terminal node represents a class label. The top node in a tree is said to be root node.

Making Tree (Training Database T)
Subdivide (T);
Building Tree (Database D)
If (each record in D is in similar class)
Return;
For every attribute A
Utilize better splitting point observed to subdivide D1 into D2;
Subdivide (D1);
Subdivide (D2);

B. Object Oriented Design for Fast Classifier Mining Algorithm

The supervised learning algorithm handles in different stages Pre-Processing, Fast Classifier Mining Algorithm, Algorithm Implementation and Best mid-point.

Pre-Processing

The method starts by analyzing the data from a huge amount of datasets and finding the number of attributes both categorical and numeric. We have utilized seven attributes (Sex, Age, Sleep, Weight, Sports, Drink, and BP) BP is the class value. It contains 5 numerical attributes and 2 categorical attributes. These data have been stored in an array is known as "attribute". The occurrence of the final value is a class value of the "attribute" list and are given table 1.

TABLE 1: ATTRIBUTE NAME

Attribute Name
Categorical {string} Numerical {integer} BP{ Low BP, Normal BP, High BP}

The numerical attributes are divided and attributes lists are arranged. Attribute lists are generated are now each attribute is sorted using the generic quick sort algorithm [5]. The sorted values in each numerical attribute and stored in array along with mid-point as a structure of value, as shown table 2

TABLE 2: SORTING ATTRIBUTES

Sorting attribute (Quick sort)
Attribute label {value} Midpoint= Total number of attribute value /2

III. PROPOSED METHOD

Input: Set of attributes having n attributes contains A= {a1, a2,…,an}from data base
Output: Construct of the DT is based on the node distribution count value.

Get back the input value from the dataset.
The categorical attribute value is converted to a numeric value. Sex to a numerical value (male-0 and female-1).
Sorting the input values utilizes the quick sort algorithm of all numeric attributes.
For every attribute develop a split point value (Illustration, age, sleep, weight, sports, and drink).

a_i is an name of attribute and vi is the splitting point values for every attributes, Initially value of *Count =0* represents class count value and *Missing=0*indicates the missing count value.
For I = 1 TO N // N implies node count in an attribute.
Search the attribute of every record.
IF $a_i \le v_i$ is true go to or move left node to traverse upto N Node
IF counting class value exist in the traveling node then manipulate the correct class count value.
Count = Count + 1;
else
IF $a_i \le v_i$ is false, then goto or move right node traverse upto *N* node.
Counting the class value in the traveling node then updates the appropriate count value of the class.
Count = *Count* + 1;
else
missing count values to be updated.
Missing = Missing + 1;
End IF
End IF
End For
Get the outcome node list class value from the database and store the appropriate count value in the table.
The construct of the decision tree is based on the distribution of the node count value.

N1	sex=M Node goto N2 else N3	N23	sleep<=6 Node goto N46 else N47	N45	drink<=4 Node goto N90 else N91
N2	age<=35 Node goto N4 else N5	N24	sleep<=6 Node goto N48 else N49	N46	drink<=4 Node goto N92 else N93
N3	age<=35 Node goto N6 else N7	N25	sleep<=6 Node goto N50 else N51	N47	drink<=4 Node goto N94 else N95
N4	weight<=48 Node gpto N8 else N9	N26	sleep<=6 Node goto N52 else N53	N48	drink<=4 Node goto N96 else N97
N5	weight<=48 Node goto N10 else N11	N27	sleep<=6 Node goto N54 else N55	N49	drink<=4 Node goto N98 else N99
N6	weight<=48 Node goto N12 else N13	N28	sleep<=6 Node goto N56 else N57	N50	drink<=4 Node goto N100 else N101
N7	weight<=48 Node goto N14 else N15	N29	sleep<=6 Node goto N58 else N59	N51	drink<=4 Node goto N102 else N103
N8	sport<=4 Node goto N16 else N17	N30	sleep<=6 Node goto N60 else N61	N52	drink<=4 Node goto N104 else N105
N9	sport<=4 Node goto N18 else N19	N31	sleep<=6 Node goto N62 else N63	N53	drink<=4 Node goto N106 else N107
N10	sport<=4 Node goto N20 else N21	N32	drink<=4 Node goto N64 else N65	N54	drink<=4 Node goto N108 else N109
N11	sport<=4 Node goto N22 else N23	N33	drink<=4 Node goto N66 else N67	N55	drink<=4 Node goto N110 else N111
N12	sport<=4 Node goto N24 else N25	N34	drink<=4 Node goto N68 else N69	N56	drink<=4 Node goto N112 else N113
N13	sport<=4 Node goto N26 else N27	N35	drink<=4 Node goto N70 else N71	N57	drink<=4 Node goto N114 else N115
N14	sport<=4 Node goto N28 else N29	N36	drink<=4 Node goto N72 else N73	N58	drink<=4 Node goto N116 else N117
N15	sport<=4 Node goto N30 else N31	N37	drink<=4 Node goto N74 else N75	N59	drink<=4 Node goto N118 else N119
N16	sleep<=6 Node goto N32 else N33	N38	drink<=4 Node goto N76 else N77	N60	drink<=4 Node goto N120 else N121
N17	sleep<=6 Node goto N34 else N35	N39	drink<=4 Node goto N78 else N79	N61	drink<=4 Node goto N122 else N123
N18	sleep<=6 Node goto N36 else N37	N40	drink<=4 Node goto N80 else N81	N62	drink<=4 Node goto N124 else N125
N19	sleep<=6 Node goto N38 else N39	N41	drink<=4 Node goto N82 else N83	N63	drink<=4 Node goto N126 else N127
N20	sleep<=6 Node goto N40 else N41	N42	drink<=4 Node goto N84 else N85	N64	Terminated 100% with H
N21	sleep<=6 Node goto N42 else N43	N43	drink<=4 Node goto N86 else N87	N65	Terminated 100% with L
N22	sleep<=6 Node goto N44 else N45	N44	drink<=4 Node goto N88 else N89	N66	Terminated 100% with H

Figure 1. Predicted Rules for Medical Dataset In Bp

After preprocessing fast classifier algorithm is initiated for the execution of object-oriented programming (OOP). Based on predicted rule classification is started by considering the split point used in BP dataset. From the database splitting point value is related to all records and a travelling path is also generated. For true condition, the node can goto left hand side node up upto n records. For false condition, splitting node goto right hand side node around and upto n records else use condition for missing count value and frequency distribution is designed.

As per proposed model, the data that has satisfied with condition, failed with condition and even the missing data count are observed. After this examination, chart is created for all distributed nodes. The DT generated may be a binary tree based on node distribution count value. Therefore, 2n-1 represents the total nodes in the tree, with n representing number of attributes used. For this condition, the attribute count is 7 (age, sex, drink, weight, BP, sleep, sports). The total nodes are 128-1=127.

IV. BEST SPLIT POINT

The splitting point is tested on every node, once split point is known and after searching all record in database, the node can be classified using IF*(x1<=v1) AND (x2<=v2) AND…………..AND(xn<=vn) THEN C rule[13].*The predicted rules is Figure1.

A. Rule Representation of the Algorithm

IF condition then class, the rule originator (IF) consists of different conditions that are associated by a logical operator (AND) [13]. The rule is as shown in Figure1.

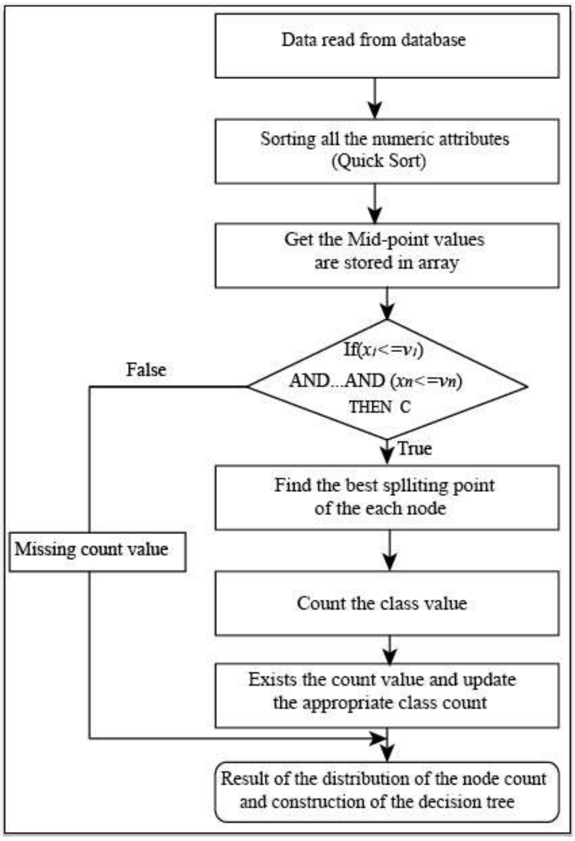

Figure 2. Design of Proposed Method

1. If (Sex == M (or) F) rule is true, then goto or move to left hand side node and if the rule is false goto or move to right node of another node age, next splitting point is based on the split-point value of age.
2. IF(Sex==M (or) F and Age ≤ (or) > split-point value) this rule is true, then goto or move to left hand side node and the rule is false goto or move right of another node weight, next splitting point is created on the split-point value of weight.
3. IF(Sex == M (or) F AND Age ≤ (or) >split-point value AND Weight ≤ (or) >split-point value) this rule is true, then goto or move to left hand node and the rule is false goto or shift to right of another node sports, next splitting point is established on the split-point value of sports.
4. IF (Sex == M (or) F AND Age ≤ (or) >split-point value AND Weight ≤ (or) >split-point value AND Sports ≤ (or) >split-point value) this rule is true, then goto or move to left hand side node and if the rule is false, then goto or move to right side of another node sleep, next split point is created on the split-point value of sleep.
5. IF (Sex == M (or) F AND Age ≤ (or) >split-point value AND Weight ≤ (or) >split-point value AND Sports ≤ (or) >split-point value AND Sleep ≤ (or) >split-point value), this rule is true goto or move to left node and the rule is false goto right of another node Drink.
6. IF(Sex == M (or) F AND Age ≤ (or) >split-point value AND Weight ≤ (or) >split-point value AND Sports ≤ (or) >split-point value AND Sleep ≤ (or) >split-point value AND Drink ≤ (or) > split-point value) this rule is true, then go to or shift to left side node and the rule is false, goto or shift to right node.
7. To end with count BP class value, if there is a presence of similar value count the current node value and update the count in the class else update the count of missing value. Node count distribution is analyzed on predicted rule (Figure.1) and histograms of the nodes that are classified are measured by constructing a DT (Figure.3, Figure.4 and Figure.5). The plan of fast classifier mining procedure is presented in Figure. 2.

B. Case Study for Medical Database

To test the usefulness of our classification algorithm, we use two state of data mining is used. Test have been agreed out to calculate the classification accuracy. The case study for medical database where the risk of having blood pressure(BP) data and node distribution count value is based on the prediction rules for medical database in BP (refer the Figure-1). The data sets containing records of the following attributes are given below.

In two different attribute BP and SEX algorithm is used. Here, SEX category can be either Male or Female and BP category as High, Low and normal of five numerical attributes are considered. age indicates a person age in year format, Weight is represented in kilogram for a person, sports highlights with exercise person in range 1-10, sleep to imply a person sleeping time in range 0-24hrs on average, drinking expels whether a person is addicted to alcohol. After classification of attributes, 24 distributes nodes are obtained with path travel from 30,000 records that was useful for counting patience with BP categories. Every node distributed are generated by IF < condition > then rule. It is produced based on prediction rule of class values, travel path and node count distribution is shown in table3.

C. Total Count Values

Sex, Weight, age, sleep, drink, sports, and BP are taken as attribute node values. From analysis out of 30,000 patients, patients with high BP are 10789, normal BP patients are 9404, low BP patients are 9470 and missing values 328. This is described as histogram of all the distribution (refer the Figure. 3 and Figure. 4)

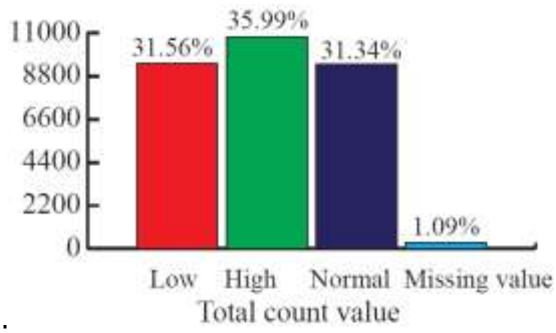

Figure 3. Total class count values

Based on classification tree is generated to classify all distributed node values and travel path using 30,000 records. (Refer the Figure. 4). This dataset has been classified from distribution of travelling path and class count values for medical database in BP.

V. EVALUATION RESULTS

The data set was created manually from medical database in BP. It enclosed 30,000 records with the attributes of sex, age, weight, sport, and sleep. For the experimental purpose, the outcomes are created as decision tree by utilizing the decision tree algorithm and fast classifier mining algorithm to count the distribution class value and travelled path for medical database in BP. The tree is divided into 24 classification nodes and also travel path. The class values are considered missing count, high BP, low BP, and normal BP.

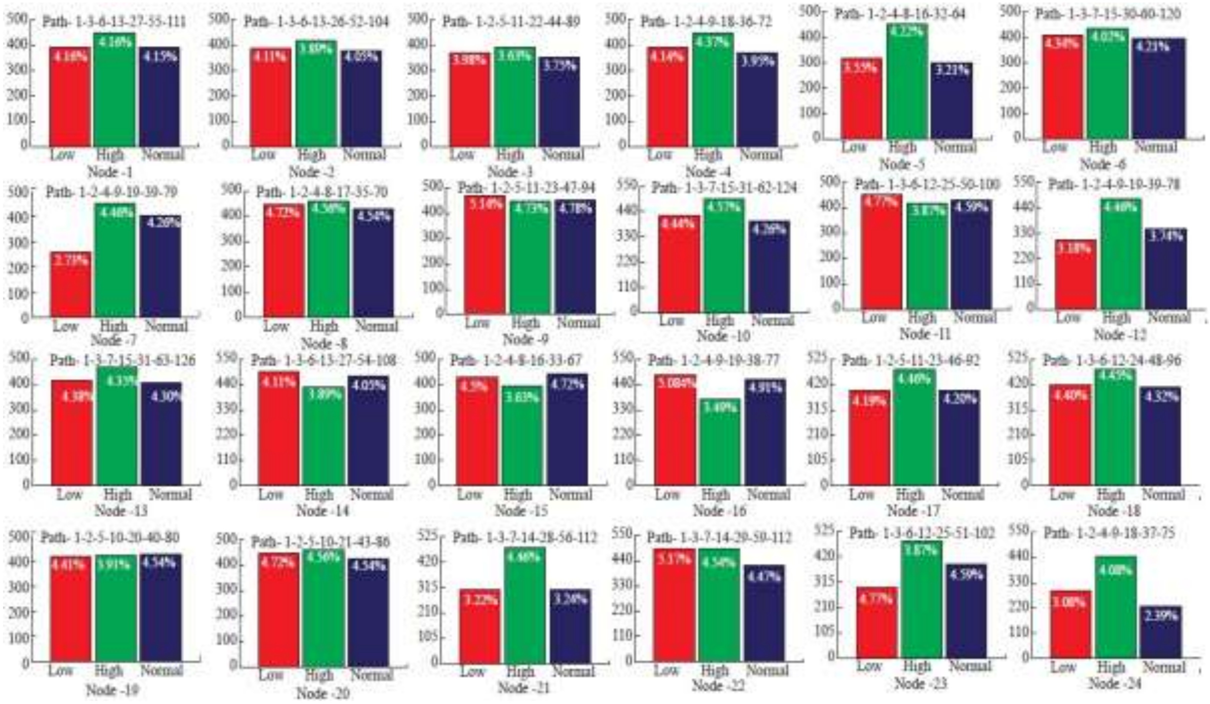

Figure 4. Distribution of the Node Count Values

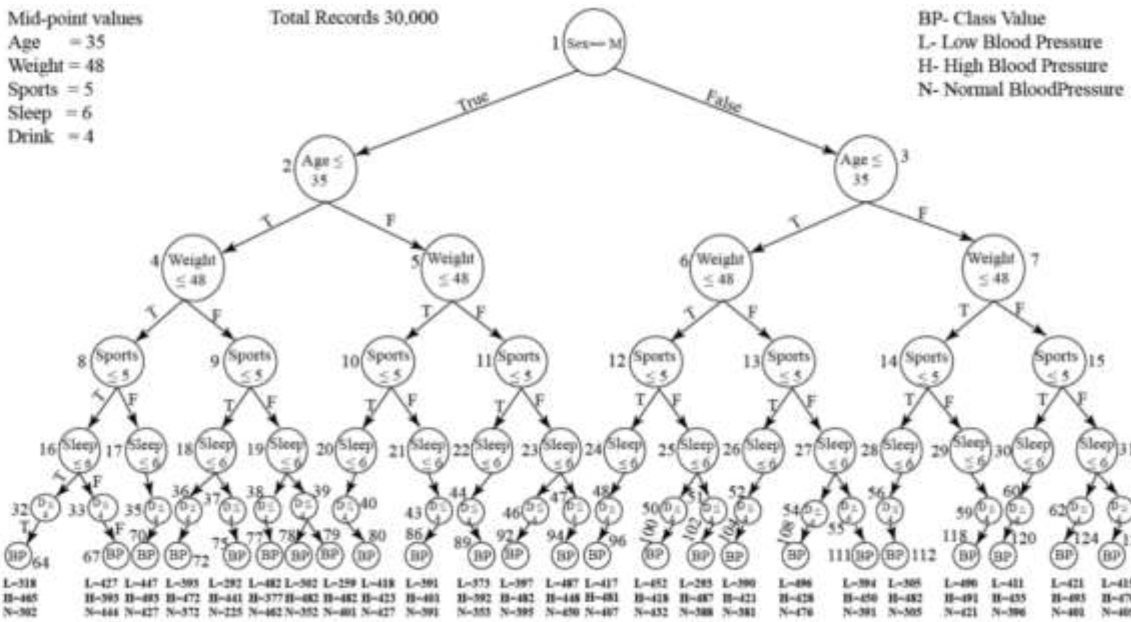

Figure 5. Classification tree

VI. EVALUATION METRICS

Data mining techniques requires a calculation process that verifies the methods produced by fast classifier algorithm. All classification models can be analyzed using labels in supervised methods by different metrics like accuracy, precision, recall and F-measure.

TABLE 3: CONFUSION MATRIX

Predicted class			
Known class	L	H	N
L	tpL	eLH	eLN
H	eHL	tpH	eHN
N	eNL	eNH	tpN

TABLE 4: EXPERIMENTAL AND RESULTS

Data sets		Confusion Matrices			Results			Accuracy %
		L	H	N	Precision %	Recall %	F-Measure %	
Dist 1	L	394	2	5	98.25	98.25	98.25	98.25
	H	3	450	2	99.34	98.90	99.12	
	N	4	1	391	98.24	98.74	98.49	
Dist 2	L	390	1	9	98.24	98.98	98.61	98.43
	H	4	421	3	98.59	98.36	98.48	
	N	3	5	381	98.45	97.94	98.20	
Dist 3	L	373	0	6	98.16	98.42	98.29	98.42
	H	5	392	1	97.51	98.49	98.00	
	N	2	4	353	99.72	98.33	99.02	
Dist 4	L	393	5	2	97.76	98.25	98.00	98.25
	H	6	472	4	98.54	97.93	98.23	
	N	3	2	373	98.42	98.68	98.55	
Dist 5	L	318	2	2	99.07	98.76	98.91	99.35
	H	1	456	0	99.56	99.78	99.67	
	N	2	0	302	99.34	99.34	99.34	
Dist 6	L	411	2	0	99.04	99.52	99.28	99.04
	H	1	435	4	99.09	98.86	98.98	
	N	3	2	396	99.00	98.75	98.88	
Dist 7	L	259	2	0	98.11	99.23	98.67	98.70
	H	4	482	5	98.97	98.17	98.57	
	N	1	3	401	98.77	99.01	98.89	
Dist 8	L	477	2	0	99.38	99.58	99.48	99.22
	H	1	493	2	98.80	99.40	99.10	
	N	2	4	427	99.53	98.61	99.07	
Dist 9	L	487	1	2	98.98	99.39	99.19	99.07
	H	3	448	3	99.33	98.68	99.01	
	N	2	2	450	98.90	99.12	99.01	
Dist 10	L	421	1	3	99.29	99.06	99.18	99.32
	H	2	493	0	99.40	99.60	99.50	
	N	1	2	401	99.26	99.26	99.26	
Dist 11	L	452	3	1	98.26	99.12	98.69	98.71
	H	5	418	1	98.35	98.58	98.47	
	N	3	4	432	99.54	98.41	98.97	
Dist 12	L	302	3	0	98.37	99.02	98.69	98.70
	H	4	482	5	98.97	98.17	98.57	
	N	1	2	352	98.60	99.15	98.88	
Dist 13	L	415	4	2	99.28	98.57	98.93	99.00
	H	1	470	3	98.95	99.16	99.05	
	N	2	1	405	98.78	99.26	99.02	
Dist 14	L	496	4	3	99.40	98.61	99.00	99.03
	H	1	428	2	98.85	99.30	99.07	
	N	2	1	405	98.78	99.26	99.02	
Dist 15	L	427	5	3	98.61	98.16	98.39	98.44
	H	3	393	4	98.25	98.25	98.25	
	N	3	2	444	98.45	98.89	98.67	
Dist 16	L	482	2	4	99.38	98.77	99.08	99.03
	H	1	377	1	98.69	99.47	99.08	
	N	2	3	462	98.93	98.93	98.93	
Dist 17	L	397	1	2	99.25	99.25	99.25	99.53
	H	1	482	0	99.79	99.79	99.79	
	N	2	0	395	99.50	99.50	99.50	
Dist	L	417	2	1	98.58	99.29	98.93	98.64

18	H	3	481	5	98.77	98.36	98.57	
	N	3	4	407	98.55	98.31	98.43	
Dist 19	L	418	7	9	99.29	96.31	97.78	98.22
	H	2	423	1	97.69	99.30	98.49	
	N	1	3	427	97.71	99.07	98.39	
Dist 20	L	391	1	1	99.49	99.49	99.49	99.41
	H	1	401	2	99.50	99.26	99.38	
	N	1	1	391	99.24	99.49	99.36	
Dist 21	L	305	2	4	98.07	98.07	98.07	98.38
	H	3	482	2	98.77	98.97	98.87	
	N	3	4	305	98.07	97.76	97.91	
Dist 22	L	490	1	2	99.39	99.39	99.39	99.36
	H	2	491	2	99.59	99.19	99.39	
	N	1	1	421	99.06	99.53	99.29	
Dist 23	L	293	1	1	99.32	99.32	99.32	99.40
	H	1	487	2	99.59	99.39	99.49	
	N	1	1	388	99.23	99.49	99.36	
Dist 24	L	292	1	0	98.98	99.66	99.32	99.27
	H	2	441	2	99.55	99.10	99.32	
	N	1	1	225	99.12	99.12	99.12	

In Table-3, confusion matrix is shown, in this all data that are classified can be represented in diagonal elements and a non-diagonal element represents incorrect classified data. Precision is calculated by the relation between true positive among both error negative and true positive.

Precision: it is a data driven process; it implies the proportion of data retrieved from the data connected set relevant to search process. It all represents the number of instances that are classified correctly in the confusion matrix (true positive- classified data that are correct, error negative-incorrect classification of data).

$$Precision = \frac{tpL}{tpL + eHL + eNL} \quad (1)$$

Where tpL indicates true positive for the class L, eHL, eNL represents error negative.

Recall: This is data retrieval, recall is used to indicate the proportion of data retrieved from connect data sets that are relevant to the query which is successful. It also represents the number of classes that are classified correctly by the proportion of true positive and both error negative and true positive [14].

$$Recall = \frac{tpL}{tpL + eLH + eLN} \quad (2)$$

Where tpL implies true positive for class L, eLH, eLN represents error negative

F-measure: This is calculated by the harmonic mean between recall and precision.

$$F\text{-}Measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

Accuracy: it is evaluated by the ratio between true negative, true positive and true result from available data.

$$Accuracy = \frac{tpL + tpH + tpN}{tpL + eLH + eLN + tpH + eHL + eHN + eNL + eNH + tpN} \quad (4)$$

The decision tree classification algorithms runs on the test data sets and processes each records. It classifies the records into low BP, high BP and normal BP. The results are verified for the above proposed classifier. These results are shown in Table-4. The evaluated results by the recall, precision, f-measure, accuracy are shown in Figure.6 and Figure.7. The proposed method is used to classify the data into classified nodes and traversed path, which is shown in Table4. The evaluated precision, recall, and F-measure are shown in Figure4. The proposed method is used to classify the data into 24 classified nodes as shown in Table4.

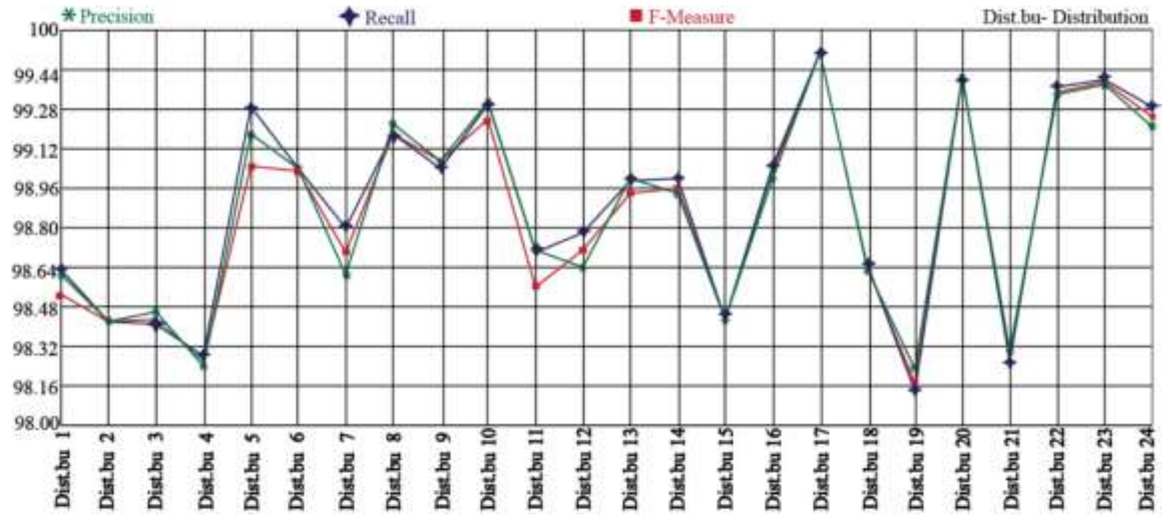

Figure 6. Scalability of Precision, recall and F-Measure

The average of the precision, recall, F-measure, and accuracy for each distribution is calculated and compared as shown in Figure.6 and Figure.7. The best result is achieved by the distribution 17 with 99.41 % precision, 99.51% recall, 99.25% F-measure and the best result for accuracy 99.53%.

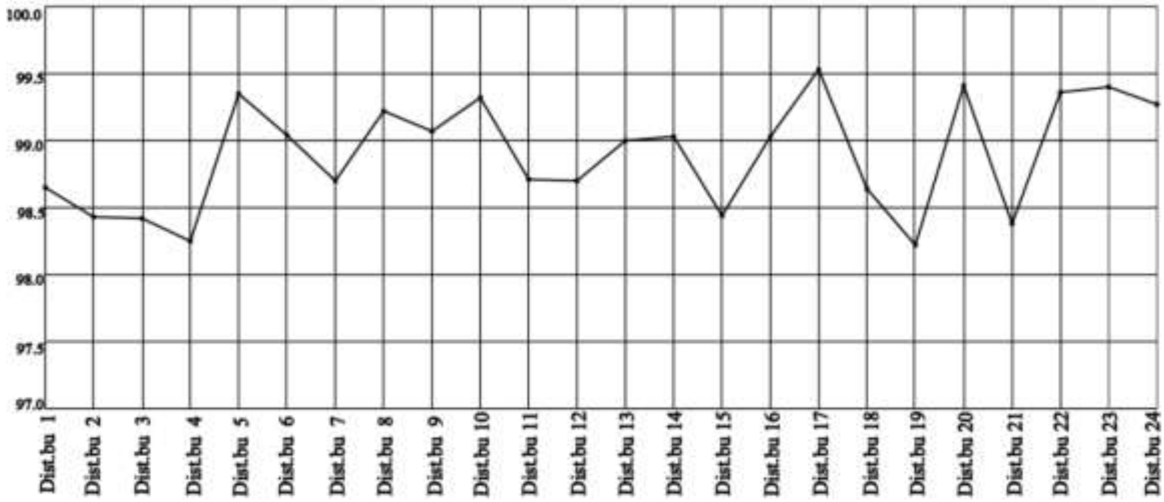

Figure 7. Scalability of Accuracy

VII. CONCLUSION

This paper has planned a new algorithm to apply in medical database for the classification of blood pressure evaluation analysis and it is based on fast classifier mining algorithm. This algorithm has been developed in java using quick sort and fast classifier mining algorithm. Also it incorporates multiple levels and multiple-split point's knowledge provided by a medical database. In this case study, multiple split points and travel path are integrated with different attributes (e.g., sex, age, weight, sports, sleep, drink, BP- is class value). Finally the class values are counted by using predicted rules in BP. When the attributes are classified we get 24 as node distribution count values and travelled path using 30,000 records.

Evaluation and analysis results are achieved classified on 24 distributions of the data collection with different attributes to explain the usefulness of the planned approach to increase the accuracy of classifier. We have reached an average accuracy of 99.53% with precision 99.41 % with recall 99.51% and F-measure 99.25%. This result proposed from the medical database is show that classification by fast classifier mining algorithm is quick and accurate with least processing time. Analysis of results using fast classifier mining algorithm shows better and increase in the accuracy for classification in real time application.

REFERENCES

1. Prasanth, Y., Sreedevi, E., Gayathri, N., & Rahul, A. S. (2017). Analysis and implementation of ensemble feature selection to improve accuracy of software defect detection model. Journal of Advanced Research in Dynamical and Control Systems, 9(18), 601-613.
2. PrasanthYalla, Venkata Naresh Mandhala, ValavalaAbhishiktha, ChitturiSaisree, KandepiManogna (2019), Machine Learning Techniques to Predict Defects by using Testing Parameters, International Journal of Recent Technology and Engineering (IJRTE), 8(4), 7829-7834.
3. Rajesh Kumar. E, K.V.S.N. Rama Rao, SoumyaRanjanNayak& Ramesh Chandra. (2020). Suicidal ideation prediction in twitter data using machine learning techniques, Journal of Interdisciplinary Mathematics, 23(1), 117-125.
4. Rajesh Kumar .E & K.V.S.N. Rama Rao. (2020). Suicide Prediction in Twitter Data using Mining Techniques: A Survey, International Conference on Intelligent Sustainable Systems (ICISS 2019), 122-131.
5. Sundar S, Srikanth D and Shanmugam M.S (2006) A new predictive classifier for improved performance in data mining: object oriented design and implementation. Proceedings of the International Conference on Industrial Mathematics. IIT Bombay, Narosa, pp 491-514
6. Mehta M,Agrawal R, Rissanen J (1996) SLIQ: a fast scalable classifier for data mining. In: Proceedings of the 5th international conference on extending database technology: advances in database technology .pp 18-32
7. Shafer CJ, Agrawal R, Mehta M (1996) SPRINT: a scalable parallel classifier for data mining. In: Proceedings of the 22th international conference on very large data bases. pp 544-555
8. Pradeepini, G., Pradeepa, G., Tejanagasri, B., & Gorrepati, S. H. (2018). Data classification and personal care management system by machine learning approach. International Journal of Engineering and Technology (UAE), 7(32), 219-223.
9. Siva Kumar Pathuri, ViswaGanesh.A, Ravi Teja.P and V.RishiChowdary. (2019). Feature-Based Opinion Mining for Amazon Product's using MLT, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(11), 4105- 4109.
10. Siva Kumar. P, N.Anbazhaghan, Shaik Razia, M Sivani, S Pravalika, A S Harshini, Prediction of cardiovascular disease using classification techniques with high accuracy, JARDC, 12(2), 1134-1139.
11. S.Sivakumar,SoumyaRanjanNayak, Ashok Kumar, S. Vidyanandini, GopinathPalai, "An empirical study of supervised learning methods for Breast Cancer Diseases", " International Journal for Light and Electron Optics ", 175, pp.105-114, 2018.
12. Sreedevi, E., &Prasanth, Y. (2017). A novel class balance ensemble classification model for application and object oriented defect database, Journal of Advanced Research in Dynamical and Control Systems, 9, 702-726.
13. Ömer AKGÖEK (2013) A rule induction algorithm for knowledge discovery and classification. Turk J Elec Eng& Comp Sci 21:1223-1241.
14. Sreedevi. E, PremaLatha. V, Sivakumar. S. (2019). A Comparative Study on New Classification Algorithm using NASA MDP Datasets for Software Defect Detection, *2nd International Conference on Intelligent Sustainable Systems (ICISS 2019),* 312-317.
15. MichileD,Spiegelhalter D J, and Taylor C C (1994) Machine Learning Neural and statistical classi_cation. Ellis Horwood
16. Agrawal R,Imielinshi T, and Swami A(1993) Database mining: A performance perspective. IEEE Transaction on knowledge and Data Engineering 5(6):914-925.
17. Wirth J and Catlett J (1988) Experiments on the cost and benefits of windowing in ID3. In 5th Int'l conference on Machine Learning
18. Farhan Hassan Khan, Saba Bashir, Usman Qamar (2014) TOM: Twitter opinion mining framework using hybrid classification scheme. Decision Support System 57:245-257.
19. Agrawal R,GhoshS,ImielinskiT,Iyer B, and Swami A (1992) An interval classifier for database mining application. In: Proc. of the VLDB Conference 560-573.