

Applications of advance ML methods in Chronic Kidney Disease classification

Dr. R.S. Walse¹¹ Assitant Professor (Computer Science),

College of Dairy Technology, Warud (Pusad), M.A.F.S.U, Nagpur, Maharashtra State, India

rajeshwalse@gmail.com

Abstract. The researcher using a classification method for the comparative study of chronic kidney patient analysis of data for Supervised with SMO – SVM and Unsupervised Hierarchical Clustering function. Now we are proposed best model by applying the Chronic kidney disease data contains 25 attributes and 400 instances including class, First, we are applying Supervised- classify – SMO- SVM- **One Attributes** – htnVs Class, Applying Cross-validation from 4 to 15 + 70% split, **Calibrator:** Logistic and **Kernel:** Polynomial, and found the result, the summary of classifier model value of ROC area for CV fold using 4 is 0.794 and the weighted average is same for all CKD, not-CKD class value is also 0.794, the accuracy of correctly classified instances 74.25 % as well as the result of a Confusion matrix is same for all Cross-Validation Folds from 5 to 15, the value of ROC area CCI is same for all CV folds from 5 to 15.

And similarly, the researcher is using Second, Step-2 by applying **two attributes (htn, appetVs Class)**, – **Two Attribute** with Supervised- classify – SMO- SVM- Applying Cross-validation folds from 4 to 15 + 70% split, **Calibrator:** Logistic and **Kernel:** Polynomial function to check further prediction and trying to increase the better accuracy as compared to the earlier model, during practical research found the result the summary of classifier model value of ROC area is by using CV folds is 4 the value of ROC Area is 0.868 and CV fold is 5 the result of ROC area is 0.844. The accuracy of the result of summary classifier model by applying CV Fold is 4 and the accuracy of correctly classified instances is 83.33 %, the researcher also checks the CCI accuracy by increasing CV fold from 5 to 15, but the result is not getting the higher accuracy, as well as the value of Confusion matrix, is same for all Cross-Validation Folds from 6 to 15, the value of ROC area CCI is same for all CV folds from 6 to 15.

Similarly, research also trying to increased accuracy and allying Third, **Step-3 (24-Full Vs Class)** – **by using Full Attribute**, Supervised- classify – SMO- SVM- Full 24 Attributes – 24 Attributes Vs Class (CKD and not-CKD), Applying Cross-validation from 4 to 15 + 70% split, **Calibrator:** Logistic and **Kernel:** Polynomial function and found the result of accuracy by the class value of ROC Area is near to value 1 i.e 0.988 with accepting the value of CV fold is 4,7,11,12,13 and the another highest accuracy value of ROC Area is 0.990 which is equal to value 1.000 with accepting the value of CV folds is 5,6,8,9,10,14,15 which is the highest accuracy, similarly the value of CCI is 98.50 % for CV fold is 4,7,11,12 & 13 & 98.75 % for CV folds is 5,6,8,9,10,14,15, the result is same for confusion matrix also by using same CV folds.

Therefore, found the best model prediction for Supervised- SMO in WEKA on the basis of three test using calibrator of logistic and kernel using polynomial function by applying Cross Validation Folds from 4 to 15 (Train on a portion of the data and test on the remainder) with 70% split, and final prediction is, increasing the no of attributes the accuracy of Correctly Classified Instances (CCI), ROC Area value and Confusion matrix value increased. Similarly research is also found to confirm predict the result of Unsupervised Hierarchical Clustering algorithm by applying full attributes also confirm prediction is increasing the number of Clusters from 2,3,4 and 5 for both Euclidean and Manhattan Function, the accuracy of result in terms of better and accurate clusters found. The adopted methodology clears the process of practical.

Keywords: Data Mining, Classification, Clustering, SMO, SVM, Calibrator, Logistic, Kernel, Polynomial, Euclidean, Manhattan, WEKA, CKD.

1 Introduction

Now the computer science techniques like optimized association rule mining techniques is using for improved Genetic Algorithms data mining and machine learning are used to study the power of

various parameters and make predictions of the based on different data sets. Bharara et al. [3] reviewed to extract for business operations using Data Mining techniques. Ariff et al. [2] studied RFID based systematic livestock health management system. Jinyin [7] performed a novel cluster center is the fast determination clustering algorithm. Dilli Arasu and

Thirumalaiselvi [1] dealt for novel imputation techniques for the effective type of predictions of kidney disease patients. ZouChuan et al. [4] performed an applied study of Guangdong provincial hospital of traditional Chinatreatment. Guangzhou and explore clustering analysis for syndrome evolution peritoneal dialysis patients. Kunwar et al. [9] studied and analyzed Chronic in terms of permanent Kidney Disease harnessing of data mining for classification techniques. AnhLuong [5] applied K- Means Approach to Clustering disease Progressions. Sabri [6] used data mining techniques for segmenting customers' information. Kumar and Lhatri [10] used WEKA is used for medical related data classification and to find early disease prediction. Khanna [10], NCBI [12] performed a study on the economics of Dialysis in India. J Nephrol [13] studied the occurrence of chronic kidney disease in India, and where are we heading? Uboltham et al. [11] performed a diagnostic study of acute kidney injury using the KDIGO guideline approach.

In this research paper, three algorithms it has been analyzed one is NB Classifier, J48, and Random Forest Decision Tree. Data cleaning in DM is used to removal of noise and inconsistent data with data integration technique with the combination of multiple types of data. To evaluate the data, we have used secondary data and it is retrieved from UCI machine learning repository [14]. Jnephrol [13], with increasing life period and the frequency of lifestyle disease, the US has seen a 30% considerable growth in the widespread presence of CKD in the last decade. Unfortunately, from India, there is no longitudinal study and limited data on the incidence of CKD. At present, the living standard of the people and the daily consumption of food are adversely affecting their health, especially their everyday living, which is increasing the number of kidney diseases in India every day. His anatomy also depended on the diet of people 40 years ago or older, but today, kidney disease is not only limited to people with diabetes or hypertension, but it has many causes. Chemical cereals, vegetables, fruits are the result of all these things, This is our daily food, and where the result is not where we are on the kidney, so the loss of kidney function slowly and then become kidney failure, such things are growing. According to reference Jnephrol [13], unfortunately, from India,

there is no longitudinal study of CKD and limited data. So because of all of the above, we have tried to analyze acute kidney disease by using Naïve Bayes, decision tree J48, and random forest algorithm unprocessed learning technique. Indeed, the purpose of our research is to use our research to analyze kidney disease or whether it can cause kidney disease in the future.

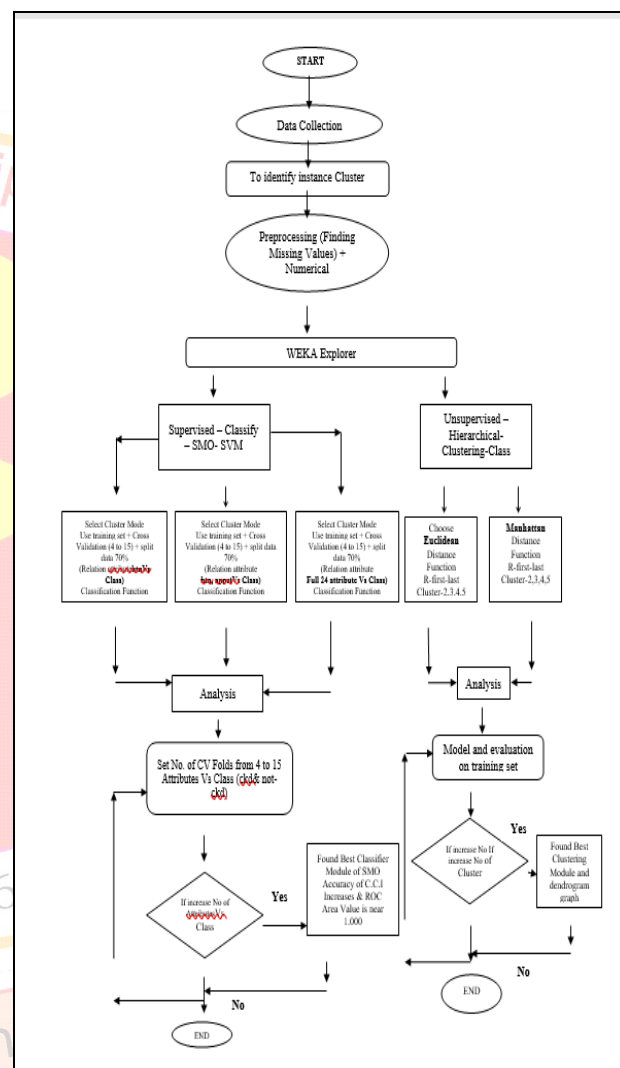


Figure 1: The research work flow

3. Result and Discussion

We are using a chronic kidney failure disease dataset, in dataset training database perfection for the Supervised SMO- SVM function and Unsupervised Hierarchical clustering techniques, and select some parameters 1. RBC count 2. Hyper-tension (BP) 3. Diabetes M. 4. Coronary disease 5. Appetite 6. Pedal Edema 7. Anemia, we are using WEKA tool for classifying and Clustering algorithms of data using Logistic and Polynomial function as well as Euclidean and Manhattan distance function.

The clinical data of kidney disease of 400 records considered for analysis has taken from the standard Machine Learning website. The data obtained after cleaning and removing missing values for further analysis, the data contains 25 attributes in the dataset with class (CKD and Not-CKD) and Class distribution is (63% for CKD and 37% for not CKD).

4. Result & Analysis

The result of experiment is to be compared of SMO Classification – SVM algorithm with Hierarchical Clustering algorithm are made established on the basis of performance in terms of high accuracy with a minimum period processing. The following algorithm is to analyze through data; the results and analysis of all three algorithms are as follows.

Supervised – Classifier – SMO-SVM- One , Two and Full Attributes Test Results using Calibrator: Logistic and Kernel : Polynomial

Step-1

Step-1 (htnVs Class) – One Attribute

Supervised- classify – SMO- SVM- 2 Attributes – htnVs Class

Applying Cross validation from 4 to 15 + 70% split

Calibrator: Logistic and **Kernel:** Polynomial

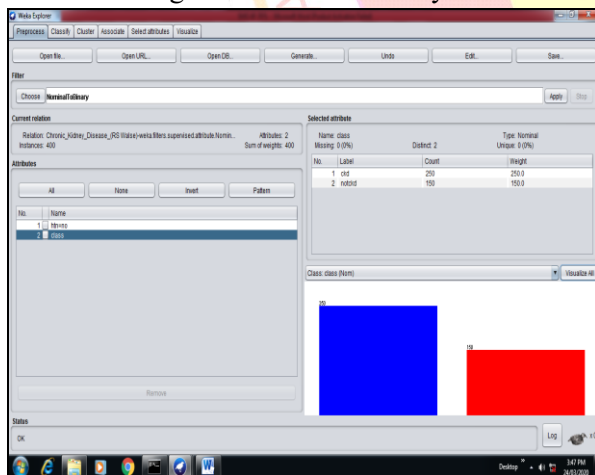


Figure: 2 Preprocess of imported data in WEKA

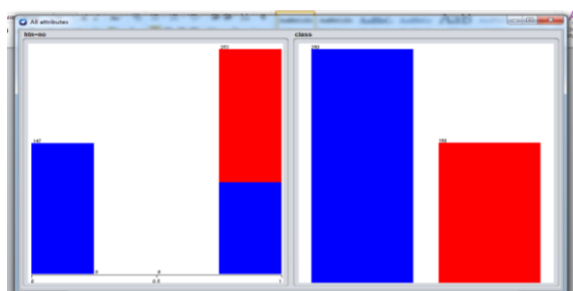


Figure: 3 Histogram of data set in WEKA

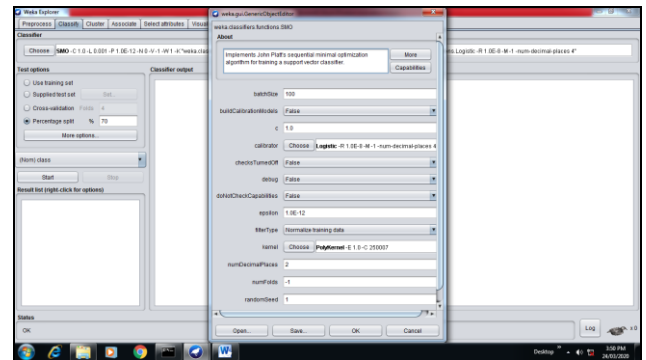


Figure: 4 SMO – applying sequential minimal optimization algorithm for training a support vector classifier

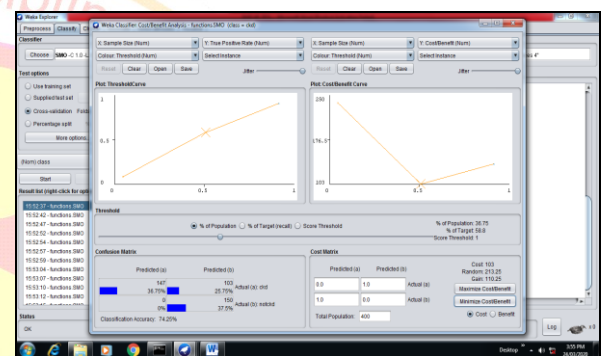


Figure: 5 Weka classifier : Cost / Benefit analysis: Function SMO (class-Ckd)

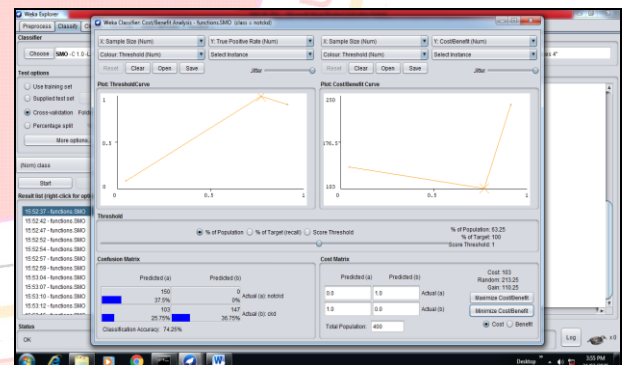


Figure: 6 Weka classifier : Cost / Benefit analysis: Function SMO (class-Not-Ckd)

Supervised – Classifier – SMO Function

Step-1

Step-1 (htnVs Class) – One Attribute

Supervised- classify – SMO- SVM- 2 Attributes – htnVs Class

Applying Cross validation from 4 to 15 + 70% split

Calibrator: Logistic and **Kernel:** Polynomial

== Classifier model (full training set) ==

SMO ,Kernel used:Linear Kernel: $K(x,y) = \langle x,y \rangle$

Classifier for classes: ckd, notckdBinarySMO,
Machine linear: showing attribute weights, not
support vectors.

2 * (normalized) htn=no, -1

**Comparative Study supervised of SMO,
calibrator: logistic and kernel: polynomial using
cross validation fold 4 to 15 with 70% of split
classifier**

Cross Validation Fold	Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
4	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
5	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
6	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
7	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
8	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
9	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
10	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
11	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
12	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
13	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
14	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
15	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight Avg	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751

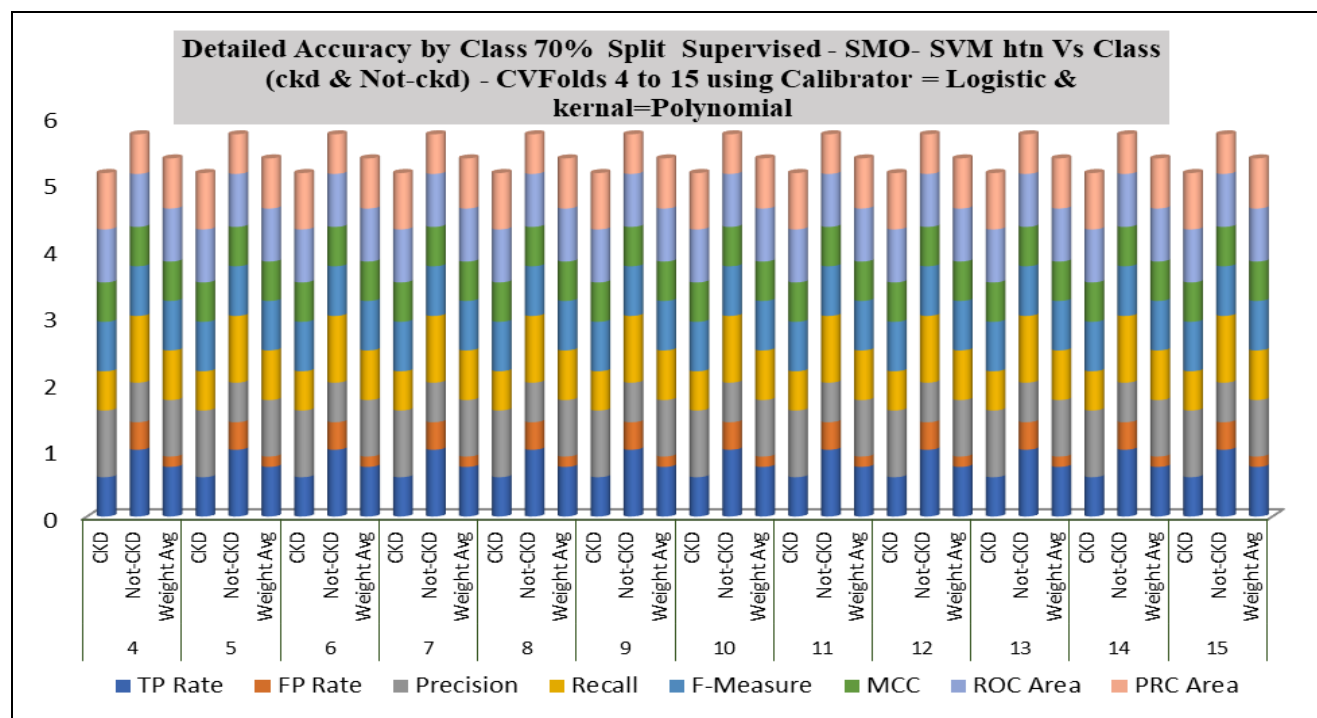


Figure: 7 detailed accuracy by class Supervised SMO-SVM attribute htn Vs class with CV folds 4 to 15 using Calibrator and Kernel function.

Table 2: Summary of Classifier model (full training set) for htn Vs Class

1. Test Mode: split 70% , 2. Total Number of Instances=400

Sr . No.	Particulars	4	5	6	7	8	9	10	11	12	13	14	15
1	Correctly Classified Instances	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25
2	Incorrectly Classified Instances	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75
3	Kappa statistic	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517
4	Mean absolute error	0.2575	0.2575	0.2575	0.2575	0.2575	0.2575	0.2575	0.2575	0.2575	0.2575	0.2575	0.2575
5	Root mean squared error	0.5074	0.5074	0.5074	0.5074	0.5074	0.5074	0.5074	0.5074	0.5074	0.5074	0.5074	0.5074
6	Relative absolute error	54.90	54.9106	54.911	54.9098	54.91	54.9108	54.9131	54.9105	54.9107	54.9101	54.9114	54.9128
7	Root relative squared error	104.81	104.817	104.8165	104.8123	104.8121	104.8128	104.8172	104.8114	104.8113	104.8098	104.8122	104.8158

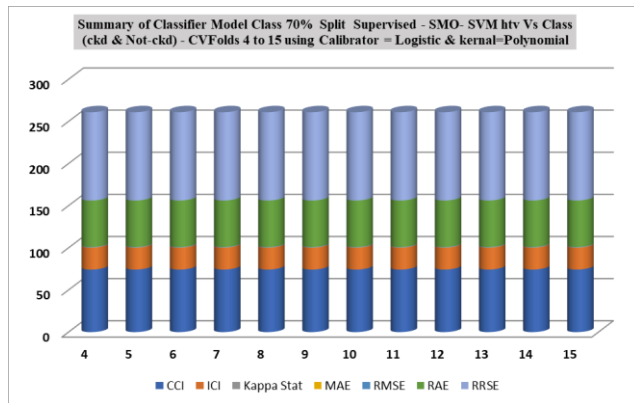


Figure: 8 Summary of classifier model class SMO-SVM attribute htnVs class with CV folds 4 to 15 using Calibrator and Kernel function.

=====**Confusion matrix**=====

The above result for all Cross Validation Folds from CVF = 4 to CVF=15 is the same

Table 3:Confusion matrix (full training set) for htnVs Class

Sr. No.	CVF	Predicted (a)	Predicted (b)	< - Classified as
1	4	147	103	a = ckd
		0	150	b = not-ckd
2	5	147	103	a = ckd
		0	150	b = not-ckd
3	6	147	103	a = ckd
		0	150	b = not-ckd
4	7	147	103	a = ckd
		0	150	b = not-ckd
5	8	147	103	a = ckd
		0	150	b = not-ckd
6	9	147	103	a = ckd
		0	150	b = not-ckd
7	10	147	103	a = ckd
		0	150	b = not-ckd
8	11	147	103	a = ckd
		0	150	b = not-ckd
9	12	147	103	a = ckd
		0	150	b = not-ckd
10	13	147	103	a = ckd
		0	150	b = not-ckd
11	14	147	103	a = ckd
		0	150	b = not-ckd
12	15	147	103	a = ckd
		0	150	b = not-ckd

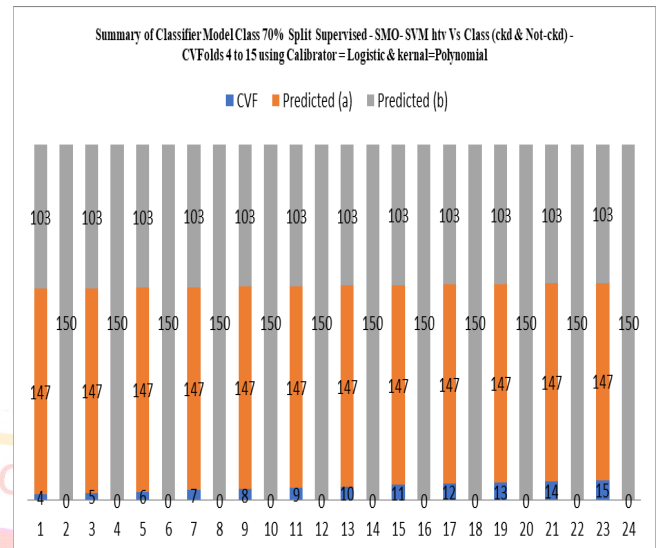


Figure: 9Confusion matrix SMO-SVM attribute htnVs class with CV folds 4 to 15 using Calibrator and Kernel function.

Step-2

Step-2 (htn, appetVs Class)

(Htn, appetVs Class) – Two Attribute

Supervised- classify – SMO- SVM- 2 Attributes –

htn, appetVs Class

Applying Cross validation from 4 to 15 + 70% split

Calibrator: Logistic and Kernel: Polynomial

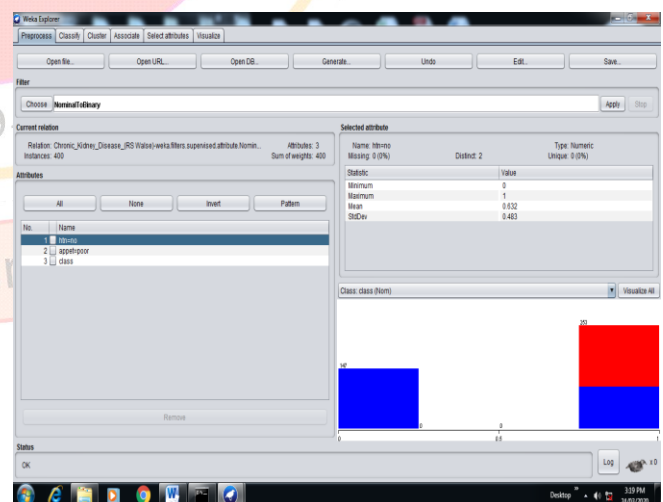


Figure: 10 Preprocess of imported data in WEKA selected two attributes

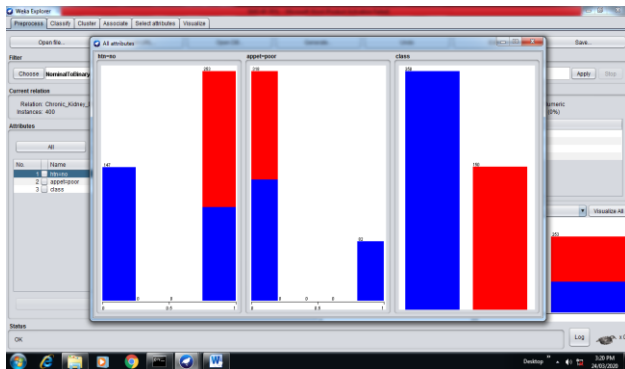


Figure: 11 Histogram of Two attributes Vs Class

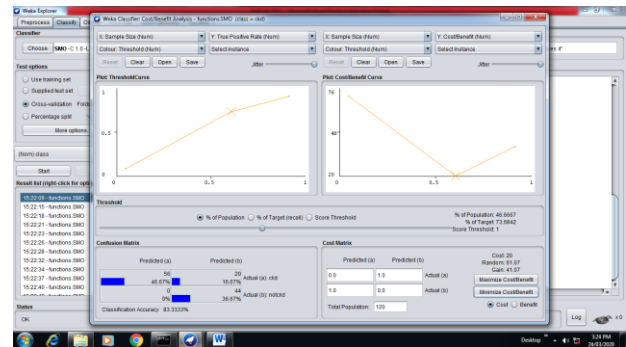


Figure: 14Weka classifier: cost / benefit analysis – function SMO class as ckd

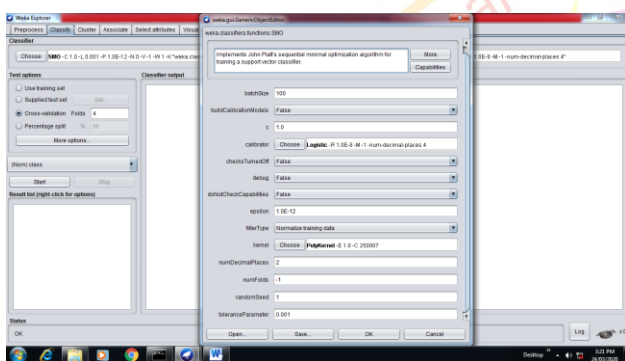


Figure: 12 SMO algorithm with properties of Logistic and Polynomial function

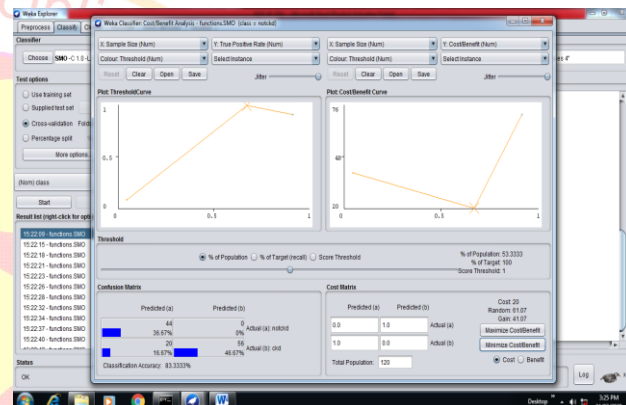


Figure: 15 Weka classifier: cost / benefit analysis – function SMO class as not-ckd

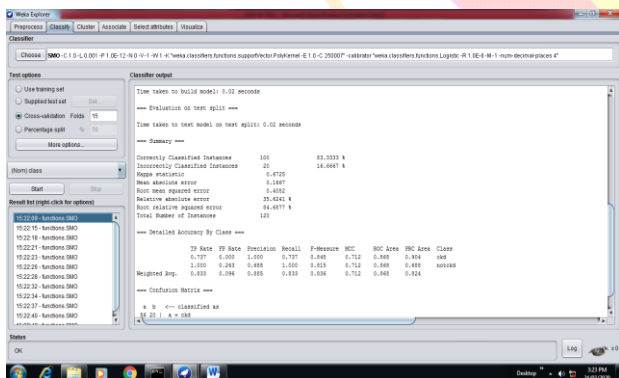


Figure: 13 Output SMO classifier with two attributes with class

Attributes: 3, htn=noappet=poorclassTest mode: split 70.0% train, remainder test

== Classifier model (full training set) ==, SMO , Kernel used: Linear Kernel: $K(x,y) = \langle x,y \rangle$ Classifier for classes: ckd, notckd ,BinarySMO , Machine linear: showing attribute weights, not support vectors.

$$2 * (\text{normalized}) \text{ htn=no} \\ + -2 * (\text{normalized}) \text{ appet=poor} \\ - 1$$

Comparative Study supervised of SMO, calibrator: logistic and kernel: polynomial using cross validation fold 4 to 15 with 70% of split classifier

Table 4: Detailed accuracy by Class with two attributes

Cross Validation Fold	Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
4	CKD	0.737	0.000	1.000	0.737	0.848	0.712	0.868	0.904
	Not-CKD	1.000	0.263	0.688	1.000	0.815	0.172	0.868	0.688
	Weight Avg	0.833	0.096	0.886	0.833	0.836	0.172	0.868	0.824
5	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.799
6	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
7	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
8	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
9	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
10	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
11	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
12	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
13	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
14	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794
15	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
	Not-CKD	1.000	0.312	0.658	1.000	0.794	0.673	0.844	0.658
	Weight Avg	0.805	0.117	0.872	0.805	0.807	0.673	0.844	0.794

Scheme:weka.clusterers.HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.ManhattanDistance -R first-last"

Relation:Chronic_Kidney_Disease_(RS Walse)-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

=== Clustering model (full training set) ===

Table 13: Comparative statement of Unsupervised Hierarchical Clustering algorithm using Euclidean and Manhattan Distance Function of full attributes with 70% split. (Cluster 2,3,4& 5)

Sr . No.	No. of Cluster	Euclidean Distance		Manhattan Distance	
		Clustered Instances	Classes to Clusters	Clustered Instances	Classes to Clusters
1	2	0 399(100%) 1 1(0%)	249 1 ckd 150 0 notckd	0 399(100%) 1 1(0%)	249 1 ckd 150 0 notckd
2	3	0 398 (100%) 1 1 (0%) 2 1 (0%)	248 1 ckd 150 0 notckd	0 398 (100%) 1 1 (0%) 2 1 (0%)	248 1 ckd 150 0 notckd
3	4	0 397 (99%) 1 1 (0%) 2 1 (0%) 3 1 (0%)	247 1 ckd 150 0 notckd	0 397 (99%) 1 1 (0%) 2 1 (0%) 3 1 (0%)	247 1 ckd 150 0 notckd
4	5	0 396 (99%) 1 1 (0%) 2 1 (0%) 3 1 (0%) 4 1 (0%)	246 1 ckd 150 0 notckd	0 396 (99%) 1 1 (0%) 2 1 (0%) 3 1 (0%) 4 1 (0%)	246 1 ckd 150 0 notckd

Table 13 describes the researcher also compare the result by using Unsupervised Hierarchical Clustering algorithm of Chronic Kidney Disease data with the support of Euclidean and Manhattan Distance clustering function with the help of 70% Split by setting the properties No of Clusters from 2,3,4, & 5 of Euclidean and Manhattan distance function. And finally research found the best and accurate prediction is the, Increasing the Number of Clusters of full attributes of Chronic Kidney Disease data accuracy of creating no of clusters as well as classes to Clusters increases. One more main comparative results also found, both the Euclidean and Manhattan distance function output result is also same for No of clusters 2,3,4& 5.

6. Conclusion

The ckd data is analyzed and predicted for diagnosed patients using data mining supervised classifiers and unsupervised clustering algorithm of SMO – SVM algorithms and Hierarchical algorithms respectively. The performance of these algorithms is compare using Weka tools. The final obtained result shows that the both the Classification and Clustering algorithm Novel discovered module is the best truthful classifier with 98.75% accuracy.i.e. Correctly Classified Instances as compared ofearlier results by applying one, two attributes by applying Calibrator: Logistic and Kernel: Polynomial function for CV folds is 5,6,8,9,10,14,15, the result is also same for confusion matrix also by using same CV folds. Therefore, found the best model prediction for Supervised- SMO in WEKA on the basis of three test using calibrator of logistic and kernel using polynomial function by applying Cross Validation Folds from 4 to 15 (Train on a portion of the data and test on the remainder) with 70% split, and final prediction is, increasing the no of attributes the accuracy of Correctly Classified Instances (CCI), ROC Area value and Confusion matrix value increased. Similarly research is also found to confirm predict the result of Unsupervised Hierarchical Clustering algorithm by applying full attributes also confirm prediction is increasing the number of Clusters from 2,3,4 and 5 for both Euclidean and Manhattan Function, the accuracy of result in terms of better and accurate clusters found. The adopted methodology clears the process of practical.For

research work, some of the attributes were measured RBC count, HP, Diabetes Mellitus, CAD, Appetite, Pedal Edema, Anemia, etc. Now future, this kind of research which will be helpful to the doctors or medical industry for prediction of CKD and not CKD patient based on their other health parameters, to minimize the growth rate of CKD patients and to control further damages of the kidney. Data mining plays an active role in predicting future kidney-related health problems. In this paper, three algorithms it has been analyzed. We have tried to analyze chronic kidney disease by using SMO- SVM classifier algorithm unprocessed learning technique with Calibrator: Logistic and Kernel: Polynomial function. Indeed, the purpose of our research is to use our research to analyze kidney disease, or whether it can cause kidney disease in the future, this will allow kidney patients who are currently in a state to know what caused the kidney disease, and those who have not had a kidney disease will see if they can develop kidney disease in the future, so they will not need to perform additional tests and save money.

7. Acknowledgement

Authors are grateful to the UCI ML repository is provided that all types of necessary database WEKA for providing such a reliable tool to extract and analyze knowledge from the database.

References

1. Arasu, S. Dilli, and R. Thirumalaiselvi. "A novel imputation method for effective prediction of coronary Kidney disease." In *2017 2nd International Conference on Computing and Communications Technologies (ICCCCT)*, pp. 127-136. IEEE, 2017.
2. Ariff, M. H., I. Ismarani, and N. Shamsuddin. "RFID based systematic livestock health management system." In *2014 IEEE Conference on Systems, Process and Control (ICSPC 2014)*, pp. 111-116. IEEE, 2014.
3. Bharara, Sanyam, A. Sai Sabitha, and Abhay Bansal. "A review on knowledge extraction for business operations using data mining." In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, pp. 512-518. IEEE, 2017.
4. Chuan, Zou, Tang Ying, Bai Li, Zeng Yuqun, and Lu Fuhua. "Application of clustering analysis to explore syndrome evolution law of peritoneal dialysis patients." In *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 23-26. IEEE, 2013.
5. Due Thanh Anh Luong, Dept of the computer. Sci&Engi, uni, at Buffalo Buffalo, NY, USA A K- Means Approach to Clustering disease Progressions IEEE Keywords, Sep 2017
6. Güllüoğlu, Sabri Serkan. "Segmenting customers with data mining techniques." In *2015 Third International Conference on Digital Information, Networking, and Wireless Communications (DINWC)*, pp. 154-159. IEEE, 2015.
7. Jinyin, Chen, et al. "A novel cluster center fast determination clustering algorithm." *Applied Soft Computing* 57 (2017): 539-555.
8. Khanna, Umesh. "The economics of dialysis in India." *Indian journal of nephrology* 19, no. 1 (2009): 1.
9. Kunwar, Veenita, et al. "Chronic Kidney Disease analysis using data mining classification techniques." *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*. IEEE, 2016.
10. Narander Kumar, Sabita KLhatri, Department of Computer, 3rd IEEE International Conference on Computational Intelligence and Communication Technology (IEEE-CICT 2017) "Implementing WEKA for medical data classification and early disease prediction 978-1-50, 2017
11. Uboltham, Issariya, Nakornthip Prompoon, and Wirichada Pan-Ngum. "AKIHelper: Acute kidney injury diagnostic tool using KDIGO guideline approach." In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1-6. IEEE, 2016.
12. Khanna, Umesh. "The economics of dialysis in India." *Indian journal of nephrology* 19, no. 1 (2009): 1.
13. Varma, P. P. "Prevalence of chronic kidney disease in India-Where are we heading?." *Indian journal of nephrology* 25, no. 3 (2015): 133.
https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease