

A Comparative Experiment Study on Supervised Classifier SMO – Support Vector Classifier and Unsupervised Hierarchical Cluster for Chronic Kidney Disease Data Statistics

Dr. Rajesh S. Walse¹, Dr. Pawan S. Wasnik², Dr. Hemant S. Mahalle³

¹Department of Dairy Business Management, College of Dairy Technology, Warud (Pusad), Animal and Fishery Sciences University, Nagpur, Maharashtra, India

²Department of Computer Science, New Model Degree College, Hingoli, S.R.T.M. University, Nanded, Maharashtra, India

³Department of Computer Science, Shri Vitthal Rukhmini College, Sawana, Tal, Mahagaon, Dist, Yavatmal, Sant Gadge Baba Amaravati University, Amaravati, Maharashtra, India

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

Copyright: [©] the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited



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 are 5,6,8,9,10,14,15, the result is same for confusion matrix also by using same CV folds.

Again researcher is also comparing the result with unsupervised, hierarchical clustering algorithm by applying full attributes i.e of 24 attributes with 400 instances, also found the best and accurate result and highest accuracy is the prediction of hierarchical algorithm model.

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.

I. INTRODUCTION

INTRODUCTION

to Now the computer science techniques likeoptimized association rule mining techniques is using for improved Genetic Algorithmsdata mining and machine learning are used to study the power of various parameters and make predictions of the based on different data sets. Data mining techniques is the process of identifying the hidden patterns from the big and tedious data. This may provide a vital role in the decision making for large data, not only agriculture but also health-related problems. Bharara et al. [3] reviewed to extractfor 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. DilliArasu 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. This paper Experiment has carried out on chronic kidney disease patient based on their relationship attributes, nowadays chronic kidney disease patient in India is increased day by day because of their eating habit and other health issues. Still, from the last ten years, CKD patient numbers it is increased tremendously Indian Journal of Nephrology et al.[12], therefore, in 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 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 CK. 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

Figure 1: Adopted methodology flowchart

II. 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 usingLogistic and Polynomial function as well as Euclidean and Manhattan distance function.

The clinical data ofkidney 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).

III. 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 performancein 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

Programmer Classify Cluster Associate Select attributes Viewalize				
epiocess Classily Cluster Associate Select attributes Visualize		~~~~~	2.7	
Open file Open URL Open DB	Gene	erate	Undo	Edit Save
r				
Choose NominalToBinary				Apply Sta
rrent relation		Selected attribute		
Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.supervised.attribute.Nomin Instances: 400	Attributes: 2 Sum of weights: 400	Name: class Missing: 0 (0%)	Distinct 2	Type: Nominal Unique: 0 (0%)
tributes		No. Label	Count	Weight
		1 ckd	250	250.0
All None Invert	Pattern	2 notckd	150	150.0
No. Name				
1 🔲 htn=no				
2 📃 class				
		Class: class (Nom)		Visuali
		250		
				150
Deserve				
Remove				
atus				
ж				Log
				Desktop * 10 + 3:47 P





W Mcrosoft Weid (Product Activation Failed)			
File Home Insert Page Layout References Mailings Review View Soda PDF 2012			۵ 🕜
Paste of Format Painter B httn=no class	hange	A Find * tab Replace	
Clipboard 75 250	6	Editing	
	100%	• • •	* 0 3 47 PM

Figure: 3Histogram of data set in WEKA

🔕 Weka Explorer	🕝 weka.gui.GenericObject	Editor	×	
Preprocess Classify Cluster Associate Select attributes Visual	weka.classifiers.functions.	SMO		
Classifier	About		-	
Choose SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.clas	Implements John Plat	t's sequential minimal optimization	More	ns.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
Test options Classifier output	algorithm for training a	a support vector classifier.	Capabilities	
O Use training set				
O Supplied test set	batchSize	100		
O Cross-validation Folds 4	buildCalibrationModels	False	×	
Percentage split % 70	c	1.0		
More options	calibrator	Choose Logistic -R 1.0E-8 -M -1 -num-	decimal-places 4	
(Nom) class	checksTurnedOff	False		
Start Stop	debug	False	×	
Result list (right-click for options)	doNotCheckCapabilities	False		
	epsilon	1.0E-12		
	fillerTune	Normaliza training data		
	interrype	Informatize training data	10	
	kernel	Choose PolyKernel -E 1.0 -C 250007		
	numDecimalPlaces	2		
	numFolds	-1		
	randomSeed	1		
Status	-		77.	,
ок	Open	Save OK	Cancel	Log 🛷 x0
📀 🤌 📋 🛛 🧔 😨	W			Desktop ** 🔺 👘 🔀 3:50 PM 24/03/2020

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



X: Sample Size (Num) Y: True Positive Rate (Num)		Long and a second s	and the second se	Local L
		X: Sample Size (Num)	Y: Cost/Benefit (Num)	
O -C 1.0 -L Colour: Threshold (Num) Select Instance	•	Colour: Threshold (Num)	Select Instance	es 4*
Reset Clear Open Save Jitter		Reset Clear Open Save	Jitter	
set Plot: ThresholdCurve		Plot: Cost/Benefit Curve		
tset		250		
tion Folds				
spire a				
re options. 0.5 -		176.5-		
			_	
		103	15	_
Dns.SMO Threshold				
ons.SMO		core Threshold	% of Population: 36.75	
ons.SMO		core infestidid	% of Target: 58.8	
ons.SMO			Score Threshold, 1	
ons.SMO Confusion Matrix		Cost Matrix		
ons.SMO Predicted (a) Predicted (b)		Predicted (a) Predicted (b)	Cost 103 Pandom: 212 25	
ons.SMO 147 103			Gain: 110.25	
ons.SMO 36.75% 25.75% Actual (a): ckd		0.0 1.0 Ad	Maximize Cost/Ben	nefit
0 150 Actual (b): notokd		1.0 0.0 Act	ual (b) Minimize Cost/Ben	nefit
0% 37.5%		Total Population: 400	● Cost ○ Be	enefit
Classification Accuracy: 74.25%				

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

Weka explorer			
Preprocess Classify Cl	Weka Classifier: Cost/Benefit Analysis - functions.SMO (class = notckd)		
Classifier	X: Sample Size (Num) Y: True Positive Rate (Num)	X: Sample Size (Num) Y: Cost/Benefit (Num)	
Choose SMO -C 1.0 -L	Colour. Threshold (Num) Select Instance	Colour: Threshold (Num) Select Instance	es 4"
	Reset Clear Open Save	Reset Clear Open Save	
est options			
 Use training set 	Plot: ThresholdCurve	Plot: Cost/Benefit Curve	
O Supplied test set	1	250	
Cross-validation Folds			
O Percentage split			
Here entione			
More options.	0.5 -	176.5-	
Nom) class			
Start		102	
sult list (right-click for opti	0 0.5 1	0 0.5 1	
15:52:37 - functions.SMO	Threshold		
15:52:47 - functions.SMO	% of Population % of Target (recall)	Score Threshold % of Population: 63.25	
15:52:52 - functions.SMO		% of Target 100	
15:52:54 - functions.SMO		Conternition of the	
15:52:57 - functions.SMO	Confusion Matrix	Cost Matrix	
15:53:04 - functions.SMO	Predicted (a) Predicted (b)	Predicted (a) Predicted (b) Random: 213 25	
15:53:07 - functions.SMO	150 0	Gain: 110.25	
15:53:10 - functions.SMO	37.5% 0% Actual (a): notckd	0.0 1.0 Actual (a) Maximize Cost/Benefit	
15:53:12 - functions.SMO	103 147 Actual (b): ckd	1.0 0.0 Actual (b) Minimize Cost/Benefit	
atus	25.75% 36.75%	Total Population: 400	
Mus	Classification Accuracy: 74.25%		
01/			

Figure: 6Wekaclassifier : 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) = <x,y>

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: POLYNOMICAL USING CROSS VALIDATION FOLD 4 TO 15 WITH 70% OF SPLIT CLASSIFIER

Table 1:Detailed accuracy by Class for htnVs class

Cross			FP						
Validati	Class	TP	Rate	Precisio	Recall	F-	MCC	ROC	PRC
on		Rate		n		Measure		Area	Area
Fold				1.000		~ - / /			
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
4	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
-	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
5	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
0	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
_	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
7	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
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	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
9	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
10	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
10	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
11	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
12	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
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
13	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
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
14	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								
	CKD	0.588	0.000	1.000	0.588	0.741	0.590	0.794	0.845
15	Not-CKD	1.000	0.412	0.593	1.000	0.744	0.590	0.794	0.593
	Weight	0.743	0.155	0.847	0.743	0.742	0.590	0.794	0.751
	Avg								



Figure: 7detailed accuracy by class Supervised SMo-SVM attributehtnVs class with CV folds 4 to 15 using Calibrator and Kernel function.

Table 2:S mmary of Classifier model (full training set) for htnVs Cla s1.Test Mode: split 70% , 2. Total Number of Instances=400

S	Particul	4	5	6	7	8	9	10	11	12	13	14	15
r. N	ar s												
ο													
•													



1	Correctl y Classifie d Instance	74.2 5	74.2 5	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25	74.25
2	Incorrec tly Classifie d Instanc es	25.7 5	25.7 5	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75	25.75
3	Kappa statistic	0.51 7	0.51 7	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517	0.517
4	Mean absolute error	0.2 57 5	0.257 5	0.257 5	0.257 5	0.257 5	0.257 5	0.257 5	0.257 5	0.257 5	0.257 5	0.257 5	0.257 5
5	Root mean squared error	0.5 07 4	0.507 4	0.507 4	0.507 4	0.507 4	0.507 4	0.507 4	0.507 4	0.507 4	0.507 4	0.507 4	0.507 4
6	Relative absolut e error	54.9 0	54.9 10 6	54.91 1	54.90 98	54.91	54.910 8	54.913 1	54.910 5	54.91 07	54.91 01	54.91 14	54.91 28
7	Root relati ve squar ed error	104 .8 1	104. 81 7	104.8 16 5	104.8 12 3	104.8 12 1	104.8 12 8	104.8 17 2	104.8 11 4	104.8 11 3	104.8 09 8	104.8 12 2	104.8 15 8





Figure: 8Summary 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
		147	103	a = ckd
1	4	0	150	b = not-ckd
		147	103	a = ckd
2	5	0	150	b = not-ckd
		147	103	a = ckd
3	6	0	150	b = not-ckd

		147	103	a = ckd
4	7	0	150	b = not-ckd
		147	103	a = ckd
5	8	0	150	b = not-ckd
		147	103	a = ckd
6	9	0	150	b = not-ckd
		147	103	a = ckd
7	10	0	150	b = not-ckd
		147	103	a = ckd
8	11	0	150	b = not-ckd
		147	103	a = ckd
9	12	0	150	b = not-ckd
		147	103	a = ckd
10	13	0	150	b = not-ckd
		147	103	a = ckd
11	14	0	150	b = not-ckd
		147	103	a = ckd
12	15	0	150	b = not-ckd



Supervised- classify – SMO- SVM- 2 Attributes – htn, appetVs Class Applying Cross validation from 4 to 15 + 70% split

Calibrator: Logistic and Kernel: Polynomial



Weka Explorer	~ ~ ~ ~ /	-	
Preprocess Classify Cluster Associate Select attributes Visualize Open file Open URL Open DB Gen	eraté) Undo	[Edit] [Save
Filter			
Choose NominalToBinary			Apply Stop
Current relation	Selected attribute		
Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.supervised.attribute.Nomin Attributes: 3 Instances: 400 Sum of weights: 400	Name: htn≃no Missing: 0 (0%)	Type: Nun Distinct: 2 Unique: 0 (0	meric 196)
Attributes	Statistic	Value	
All None Invert Pattern	Minimum Maximum Mean StdDev	0 1 0.632 0.483	
No. Name 1 hin-no 2 appet=poor 3 class			
	Class: class (Nom)		Visualize Al
	147		253
Remove		D	
Status	0	0.5	
ОК			Log 🛷 X
🚳 🧀 门 🔉 🌖 🚾 📼 🥥		Desktop »	▲ (*) 13:19 PM 24/03/2020

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



Figure: 11Histogram of Two attributes Vs Class



😮 Weka Explorer	weka.gui.GenericObject	ditor	×	
Preprocess Classify Cluster Associate Select attributes Visua	weka classifiers functions	SMO		
Classifier				
Choose SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.cla	Implements John Plat training a support vect	t's sequential minimal optimization algorithm for Mon or classifier.	ities	1.0E-8 -M -1 -num-decimal-places 4*
Test options Classifier output				
O Use training set	batchSize	100		
O Supplied test set		(and the second s		
Cross-validation Folds 4	buildCalibrationModels	False		
O Percentage split % 66	c	1.0		
More options	calibrator	Choose Logistic -R 1.0E-8 -M -1 -num-decimal-places 4		
(Nom) class	checksTurnedOff	False		
Start Stop	debug	False		
Result list (right-click for options)	doNotCheckCapabilities	False		
	epsilon	1.0E-12		
	filterType	Normalize training data		
	kernel	Choose PolyKernel -E 1.0 -C 250007		
	numDecimalPlaces	2		
	numFolds	4		
	randomSeed	1		
Status	toleranceParameter	0.001		
ок	Open	Save OK Car	icel	Log 🛷 x0
🚳 🖉 📋 🛛 🌖 🖭 🕢	W			Desktop * 🔹 🌖 🔭 3:21 PM

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

🔾 Weka Explorer				-	-	And Personnel						
Preprocess Classify Cluster Assoc	ciate Select attributes V	/isualize										
Classifier												
Change The office and pros					abdeamal	F 4 0 0 06000	71 collins				F.O.M. d. annual adapted allocate diff.	1
SMO -C 1.0 -L 0.001 -P 1.0E	:-12 -N U -V -1 -W 1 -K "weka	a.classmers.	functions.si	upportvector.P	olykernel -	E 1.0 -C 25000	/"-calibra	tor "weka.clas	sifiers.tunctio	ns.Logistic -R 1.L	E-8-M-1 -num-decimal-places 4"	
Test options	Classifier output											
O Use training set	Time taken to b	uild model	L: 0.02 st	econds								A.
O Supplied test set Set_												
Cross-validation Folds 15	=== Evaluation	on test sp	plit ===									
O Percentage colit % 70	Time taken to t	est model	on test :	split: 0.02	seconds							
C Percentage spint % 70	Sumary and											
More options)											
	Correctly Class	ified Inst	tances	100		83.3333	8					
(Nom) class	Kappa statistic	ssified if	istances	0.6	725	10.000/						
	Mean absolute e	rror		0.1	667							
Star Stop	Root mean squar	ed error		0.4	082							
Result list (right-click for options)	Root relative s	quared er:	ror	84.6	877 %							5
15:22:09 - functions.SMO	Total Number of	Instances	3	120								
15:22:15 - functions.SMO	Detailed Ac	curacy By	Class ===									
15:22:18 - functions.SMO												
15:22:21 - functions.SMO		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class		
15:22:23 - functions.SMO		0.737	0.000	1.000	0.737	0.848	0.712	0.868	0.904	cka		
15:22:26 - functions SMO	Weighted Avg.	0.833	0.096	0.885	0.833	0.836	0.712	0.868	0.824			
15:22:32 - functions.SMO		Colored States										
15:22:34 - functions.SMO	=== Confusion M	atrix ===										
15:22:37 - functions.SMO	a b < cl	assified a	3.5									
15:22:40 - functions.SMO	56 20 a = c	kd.										۲
Lacona en el como -												
Status												
ок												Log XC
			74								Desktop 🔭 🔺 🌒) 🐴 3:23 PM
												24/03/2020

Figure: 13Output SMO classifier with two attributes with class



Weka Explorer	Weka Classifier: Cost/Benefit Analysis - functions.SMO (class = ckd)		
lassifier	X: Sample Size (Num) Y: True Positive Rate (Num)	X: Sample Size (Num)	
Choose SMO -C 1.0 -L	Colour: Threshold (Num) Select Instance	Colour: Threshold (Num)	es 4"
est options	Reset Clear Open Save Jitter	Reset Clear Open Save Jitter	
O Use training set	Plot: ThresholdCurve	Plot: Cost/Benefit Curve	1
O Supplied test set	1	76	1 1
Cross-validation Folds			
O Percentage split			
More options.	0.5 7	48-	
Nom) class			
Start			
esult list (right-click for opti	0 0.5 1		
15:22:09 - functions SMO			
15:22:15 - functions.SMO	Threshold		
15:22:18 - functions.SMO	● % of Population ○ % of Target (recall) ○	Score Threshold % of Population: 46.6667 % of Target 73.6842	
15:22:21 - functions.SMO 15:22:23 - functions.SMO		Score Threshold: 1	
15:22:26 - functions.SMO	Confusion Matrix	Cost Matrix	
15:22:28 - functions.SMO 15:22:32 - functions.SMO	Predicted (a) Predicted (b)	Predicted (a) Predicted (b) Cost: 20	
15:22:34 - functions.SMO	56 20	Gain: 41.07	
15:22:37 - functions.SMO	46.67% 16.67% Actual (a): ckd	Actual (a) Maximize Cost/Benefit	
45:00:40 Augustan 010	0 44 0% 36 67% Actual (b): notckd	1.0 0.0 Actual (b) Minimize Cost/Benefit	1.
tatus	Classification Accuracy: 83.3333%	Total Population: 120 Cost O Benefit	
ок		Л	Log 🛷
			3:24 PM
🦉 🌽 🔚		Desktop	 1) 24/03/2020

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

Weka Explorer			- 0 X
Preprocess Classify Cl	Weka Classifier: Cost/Benefit Analysis - functions.SMO (class = notckd)		
Classifier	X: Sample Size (Num) Y: True Positive Rate (Num)	X: Sample Size (Num) Y: Cost/Benefit (Num)	
Choose SMO -C 1.0 -L	Colour: Threshold (Num) Select Instance	Colour: Threshold (Num)	es 4"
Test options	Reset Clear Open Save Jitter	Reset Clear Open Save Jitter	
O Use training set	Plot: ThresholdCurve	Plot: Cost/Benefit Curve	
O Supplied test set		76	n.
Cross-validation Folds		/	
O Percentage split 9			
More options.	0.5-7	48-	
	0.0		
(Nom) class			
Start			
Result list (right-click for opti	0 0.5 1	0 0.5 1	
15:22:09 - functions SMO	l		
15:22:15 - functions.SMO	Threshold		
15:22:18 - functions.SMO	% of Population % of Target (recall)	Score Threshold % of Population: 53.3333 % of Target: 100	
15:22:23 - functions.SMO	Q	Score Threshold: 1	
15:22:26 - functions.SMO	Confusion Matrix	Cost Matrix	
15:22:28 - functions.SMO 15:22:32 - functions.SMO	Predicted (a) Predicted (b)	Predicted (a) Predicted (b) Cost 20	
15:22:34 - functions.SMO	44 0	Gain: 41.07	
15:22:37 - functions.SMO 15:22:40 - functions SMO	36.67% 0% Actual (a): notckd	Maximize Cost/Benefit	*
45-00-40 Augetting Olio	20 56 16.67% 46.67% Actual (b): ckd	1.0 0.0 Actual (b) Minimize Cost/Benefit	
Status	Classification Accuracy: 83.3333%	Total Population: 120 Cost Benefit	
ок			Log 🛷 x0
		s it s	4. Rm 3:25 PM
🥥 🌔 🔚		Desktop	• • • • • • • • • • • • • • • • • • •

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

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) = <x,y> Classifier for classes: ckd, notckd, BinarySMO, Machine linear: showing attribute weights, not support vectors. 2 * (normalized) htn=no



+ -2 * (normalized) appet=poor - 1

COMPARATIVE STUDY SUPERVISED OF SMO, CALIBRATOR: LOGISTIC AND KERNEL: POLYNOMICAL USING CROSS VALIDATION FOLD 4 TO 15 WITH 70% OF SPLIT CLASSIFIER

Table 4:Detailed accuracy by Class with two attributes

Cross									
Validation	Class	TP	FP	Precision	Recall	F-	MCC	ROC	PRC
Fold		Rate	Rate			Measure		Area	Area
	CKD	0.737	0.000	1.000	0.737	0.848	0.712	0.868	0.904
4	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
5	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
6	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
7	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
8	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
9	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
10	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
11	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
12	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
13	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
14	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
	CKD	0.688	0.000	1.000	0.688	0.815	0.673	0.844	0.883
L									



15	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



Figure: 16Detailed accuracy by class with two attributes, CVFolds 4 to 15

Table 5:Summary of Classifier model (full	training set) with two attributes
---	-----------------------------------

S r N o	Parti cular s	4	5	6	7	8	9	1 0	11	12	13	14	1 5
1	Time	0.0	0.	0.0	0.	0	0.02		0.03	0.01	0	0.02	0
	take	2	0	1	0								
	n to		3		3								
	test												
	mod												
	el												

	Corre													
2	ctly	83.	8	80.	8	80.	80.5	80	80.50	80.5	80.5	80.5	8	
	Classi	33	0.	50	0.	50	0	.5		0	0	0	0	
	fied		5		5			0						
	Insta		0		0								5	
	nces												0	
3	Incor	16.	1	19.	1	19.	19.5	19	19.5	19.5	19.5	19.5	19	
	rectly	66	9.	5	9.	5		.5					.5	
			5		5									
	Classi		0											
	fied													
	Insta													
	nces													
4	Карр	0.6	0.6	0.6	0.6	0.6	0.62	0.6	0.623	0.62	0.62	0.62	0	
	а	725	232	23	23	232	32	23	2	32	32	32	•	
	statist			2	2			2					6	
	ic												2	
													3	
						0.1			0.107				2	
5	Mean	0.1	0.1	0.1	0.1	0.1	0.19	0.1	0.195	0.19	м_0.19	0.19	0	
	absol	007	95	95	95	95	5	95		5	5	5	•	
	ute													
	enor												5	
6	Root	0.4	0.4	0.4	0.4	0.4	0.44	0.4	0.441	0.44	0.44	0.44	0	
-	mean	082	416	41	41	416	16	41	6	16	16	16		
	squar			6	6			6					4	
	ed												4	
	error												1	
													6	
7	Relati	35.	41.	41.	41.	41.	41.5	41.	41.58	41.58	41.5	41.5	4	
	ve	624	582	583	582	582	83	58	27	29	824	834	1.	
	absol	1	8	1	2	4		47					5	
	ute												8	
	error												4	
0	Dest	0 /	01	01	01	01	01.2	01	01.20	01.20	01.2	01.2)	
ð	K00t	04. 697	ע או. זיר	ען. 212	91. 200	200 200	91.2	91. 91	91.20	91.20	91.2	91.2	у 1	
		00/ 7	213	212	6	209 5	002	21	07	00	0/5	090	1. 7	
	sallar		7	5		, J		59					1	
	ed												2	
	error												7	



9	Total	120	40	4	40	400	400	40	400	400	400	400	4
	No.		0	0	0			0					0
	of			0									0
	Insta												
	nces												



Figure: 17Summary of Classifier Model with two attributes, CVFolds 4 to 15

===== CONFUSION MATRIX ======

Table 6:Values of Confusion matrix with two attributes

Sr. No.	CVF	Predicted (a)	Predicted (b)	< - Classified as
1	4	147	103	a = ckd
		0	150	b = not-ckd
2	5	172	78	a = ckd
2	5	0	150	b = not-ckd
3	6	172	78	a = ckd
0	0	0	150	b = not-ckd
4	7	172	78	a = ckd
1	,	0	150	b = not-ckd
5	8	172	78	a = ckd
5	0	0	150	b = not-ckd
6	9	172	78	a = ckd
	,	0	150	b = not-ckd
7	10	172	78	a = ckd

I.	Ì			I
		0	150	b = not-ckd
8	11	172	78	a = ckd
0		0	150	b = not-ckd
9	12	172	78	a = ckd
,	12	0	150	b = not-ckd
10	13	172	78	a = ckd
10	10	0	150	b = not-ckd
11	14	172	78	a = ckd
11	14	0	150	b = not-ckd
12	15	172	78	a = ckd
12	15	0	150	b = not-ckd

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



Figure: 18Summary of Classifier Model with two attributes, CVFolds 4 to 15 Step-3

Step-3 (24-Full Vs Class) – 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



eprocess Classify Cluster Associate Select attribut	les Visualize			
Open file Open URL	Open DB Gene	rateU	ndo E	dit Save
r				
📄 weka				Apply Sto
* 🚔 filters		Selected attribute		
AllFilter MultiFilter RenameRelation	supervised.attribute.Nomin Attributes: 25 Sum of weights: 400	Name: class Missing: 0 (0%)	Distinct 2	Type: Nominal Unique: 0 (0%)
* 🚔 supervised		No. Label	Count	Weight
AddClassification		1 ckd	250	250.0
AttriotteSelection ClassConditionalProbabilities ClassConditionalProbabilities Discretize MergeNominalVatues NominalToBinary PartitionNembership				
 ▶		Class: class (Nom)		Visualiz
Eilter Remove filter Qiose				50
				Log

Figure: 19Preprocess of data filter attributes with Nominal to binary full attributes

	Associate Select attributes	Visualize				
Open file	Open URL	Open DB Gene	rate	Undo	Edit	Save
= weka						Apply St
filters						
- AllFilter	-)	Selected attribute			
MultiFilter		Attributes: 25	Name: age		Туре	: Numeric
RenameRelation		Sum of weights: 400	Missing: 0 (0%)	Distinct 77	Unique	: 16 (4%)
* 💼 supervised			Statistic		Value	
 attribute AddClassifical 	ion		Minimum		2	
AttributeSelect	ion	Invert Pattern	Maximum		90	
ClassConditio	nalProbabilities		StdDev		16 975	
PartitionMemb	ari					
PartitionMemb	An attribute with k values is tran option '-A' is not given. If the cla	isformed into k binary attributes if the class is nominal i ss is numeric, k - 1 new binary attributes are generated glue associated with each attribute value into account	using the one-attribute-per-value in the manner described in "Cla	e approach). Binary attributes ssification and Regression	s are left binary if Trees" by Breiman et al. (i.	Visua
Nominal robin Nominal robin PartitionMemb instance image: State of the state of th	An attribute with k values is tran option "-A" is not given. If the cla e., by taking the average class	sformed into k binary attributes if the class is nominal ss is numeric, k - 1 new binary attributes are generated value associated with each attribute value into account)	using the one-attribute-per-value in the manner described in "Cla	a approach). Binary attributes ssification and Regression	s are left binary if Trees" by Breiman et al. (i.	▼ Visua
 Nominal OBIN PartitionMemb ► instance ► instance 	An attribute with k values is tran option "A" is not given. If the da e., by taking the average class For more information, see:	sformed into k binary attributes if the class is nominal ss is numeric, k - 1 new binary attributes are generated value associated with each attribute value into accounty	using the one-attribute-per-value in the manner described in "Cla	e approach). Binary attributes ssification and Regression	s are left binary if Trees" by Breiman et al. (i.	Visua
Nominal oBin PartitionMemb ► instance ► insupervised	A contents an Annual sandcas option "A" is not given. If the class e, by taking the average class of For more information, see: L Breiman, J.H. Friedman, R.A.	sformed into k binary attributes if the class is nominal i ss is numeric, k - 1 new binary attributes are generated value associated with each attribute value into accounty Olshen, C.J. Stone (1984). Classification and Regress	using the one-attribute-per-value in the manner described in "Cla 	e approach). Binary attributes ssification and Regression	s are left binary if Trees" by Breiman et al. (i.	Visua
 Normall Osin PartitionMemb ▶ a instance ▶ a unsupervised 	A attribute with k values is tran- option f.vit is not given. If the cla e., by taking the average class to For more information, see: L. Breiman, J.H. Friedman, R.A CAPABILITIES	sformed into k binary attributes if the class is nominal i ss is numeric, k - 1 new binary attributes are generated value associated with each attribute value into account) Olshen, C.J. Stone (1984). Classification and Regress	using the one-attribute-per-value in the manner described in "Cla sion Trees. Wadsworth Inc.	e approach). Binary attributes ssification and Regression	s are left binary if Trees" by Breiman et al. (i.	Visua
 Nomnall Osin PartitionMemb instance unsupervised 	A attribute with k values is tra- option 'A' is not given. If the cla- e., by taking the average class of For more information, see: L. Breiman, J.H. Friedman, R.A. CAPABILITIES Class – Binary class, Date class	sformed into k binary attributes if the class is nominal i s is numeric, k - 1 new binary attributes are generated value associated with each attribute value into account) Olshen, C.J. Stone (1984). Classification and Regress is, Missing class values, Nominal class, Numeric class	using the one-attribute-per-value in the manner described in "Cla sion Trees. Wadsworth Inc.	e approach). Binary aftributes ssification and Regression	s are left binary if Trees" by Breiman et al. (i.	visua
Normail Osin PatitionMemb ► (in Instance ► (in unsupervised)	A controls an information subcles and a stribute with k values is tra- option 'A' is not given. If the da e, by taking the average class i For more information, see: L Breiman, J.H. Friedman, R.A CAPABILITIES Class – Binary class, Date class Attributes – Binary class, Date class	sformed into k binary attributes if the class is nominal i sis numeric, k - 1 new binary attributes are generated value associated with each attribute value into account) Olshen, C.J. Stone (1984). Classification and Regress is, Missing class values, Nominal class, Numeric class ate attributes. Errorb nominal attributes Mission values	using the one-attribute-per-value in the manner described in "Cla sion Trees. Wadsworth Inc.	e approach). Binary aftributes ssification and Regression ssification and Regression	s are left binary if Trees" by Breiman et al. (i. 5. Strinn attributes Inan	v) Visua
Nomnall Osin PartitionMemb ► @ Instance ► @ unsupervised	A antibute with k values is tra- den attribute with k values is tra- option 'A' is not given. If the da e, by taking the average class i For more information, see: L Breiman, J.H. Friedman, R.A CAPABILITES Class – Binary class, Date class Attributes – Binary attributes, D attributes	sformed into k binary attributes if the class is nominal i sis numeric, k - 1 new binary attributes are generated value associated with each attribute value into account) Olshen, C.J. Stone (1984). Classification and Regress is, Missing class values, Nominal class, Numeric class ate attributes, Empty nominal attributes, Missing values	using the one-attribute-per-value in the manner described in "Cla sion Trees. Wadsworth Inc. s , Nominal attributes, Numeric att	a approach) Binary attributes ssification and Regression ssification and Regression bributes, Relational attributes	s are left binary if Trees" by Breiman et al. (i. s, String attributes, Unary	Visua 55
Normal Tobin PartitionMemb ► @ Instance ► @ unsupervised	A attribute with k values is the option 'A' is not given. If the option 'A' is not given. If the e, by taking the average class is For more information, see: L. Breiman, J.H. Friedman, R.A. CAPABILITIES Class Binary class, Date class Attributes Binary attributes, D attributes Binary attributes Binary a	sformed into k binary attributes if the class is nominal i sis numeric, k - 1 new binary attributes are generated value associated with each attribute value into accounty Olshen, C.J. Stone (1984). Classification and Regress s, Missing class values, Nominal class, Numeric class ate attributes, Empty nominal attributes, Missing values	using the one-attribute-per-value in the manner described in "Cla sion Trees. Wadsworth Inc. s , Nominal attributes, Numeric att	e approach). Binary attributes ssification and Regression ssification and Regression tributes, Relational attributes	s are left binary if Trees" by Breiman et al. (i. s, String attributes, Unary	5 18 2
Normal10ein PartitionMemb ► () Instance ► () unsupervised	A attribute with k values is tran- option 'A' is not given. If the cla- e, by taking the average class of For more information, see: L Breiman, J.H. Friedman, R.A CAPABILITIES Class – Binary class, Date class Attributes – Binary attributes, D attributes – SupervisedFilter, V	ssormed into k binary attributes if the class is nominal i sis numeric, k - 1 new binary attributes are generated value associated with each attribute value into accounty Olshen, C.J. Stone (1984). Classification and Regress s, Missing class values, Nominal class, Numeric class ate attributes, Empty nominal attributes, Missing values veightedAttributesHandler, WeightedInstancesHandler	using the one-attribute-per-value in the manner described in "Cla sion Trees. Wadsworth Inc. s , Nominal attributes, Numeric att	e approach). Binary attributes ssification and Regression ssification and Regression tributes, Relational attributes	s are left binary if Trees" by Breiman et al. (i. s, String attributes, Unary	592
PartitionMemb PartitionMemb PartitionMemb	A attribute with k values is the a option 'A' is no typien. If the a e, by taking the average class to For more information, see: L. Breiman, J.H. Friedman, R.A CAPABILITIES Class – Binary class, Date class Attributes – Binary attributes, D attributes – Interfaces – SupervisedFilter, V Additional	sformed into k binary attributes if the class is nominal i sis numeric, k - 1 new binary attributes are generated value associated with each attribute value into account) Olshen, C.J. Stone (1984). Classification and Regress is, Missing class values, Nominal class, Numeric class ate attributes, Empty nominal attributes, Missing values VeightedAttributesHandler, WeightedInstancesHandler	using the one-attribute-per-value in the manner described in "Cla sion Trees. Wadsworth Inc. s , Nominal attributes, Numeric att	a approach) Binary attributes ssification and Regression stributes, Relational attributes	s are left binary if Trees" by Breiman et al. (i. s, String attributes, Unary	Visual

Figure: 20Convert all nominal attributes into binary numeric attributes



Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize	Constant Sectors (Sector		
Open file Open URL Open DB Gene	erate Un	do E	Edit
Filter			
Choose NominalToBinary			Apply Stop
Current relation	Selected attribute		
Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.supervised.attribute.Nomin Attributes: 25 Instances: 400 Sum of weights: 400	Name: class Missing: 0 (0%)	Distinct 2	Type: Nominal Unique: 0 (0%)
Attributes	No. Label	Count	Weight
	1 ckd	250	250.0
All None Invert Pattern	2 NOICKU	150	150.0
No. Name			
10 bgr			
11 bu			
12 SC			
13 S00			
15 hemo	Class: class (Nom)		Visualize
16 pcv			
17 wbcc			
18 Trbcc	250		
19 htn=no			
20 dm=no			
21 cad=ves			
22 appet=poor			
23 perves	· · · · · · · · · · · · · · · · · · ·		150
24 ane=ves	ckd [250]		
25 dass			
Remove			
Status OK			Log 🛷
			» 0.45 DM
🍪 🌽 📄 🔍 🦁 💾 🖆 🕢			Desktop 🔺 🕩 🔀 23/03/202

Figure: 21Convert all nominal attributes in to binary numeric attributes

Weka Explore	Classify	Cluster	Associate	Select attributes	Visualize		and suggest		Street Street								
lassifier	Classily	Cluster	Associate	Select attributes	visualize												
Choose	SMO - C 1.	0 -L 0.001	-P 1.0E-12 -	N 0 -V -1 -W 1 -K"v	/eka.classifiers	s.functions.s	upportVector.P	olyKernel -	E 1.0 -C 25000	17" -calibra	tor "weka.clas:	sifiers.functio	ns.Logistic -R 1	.0E-8 -M -1 -nui	n-decimal-plac	es 4"	
est options				Classifier output													
🔿 Use train	ning set			Time taken t	o build mode	el: 0.03 se	econda										P
 Supplied Cross-val 	I test set	Set.		=== Stratifi	ed cross-val	lidation ==											[
Percentar	ine split	96 70		O'dhand L' J													
	More optio	ons		Correctly Cl Incorrectly Kappa statis	assified Ins Classified 1 tic	stances Instances	394 6 0.9	683	98.5 1.5	*							
lom) class			•	Mean absolut Root mean sq Relative abs	e error uared error olute error		0.0	15 225 985 %									
Start		Sto	p.	Root relativ Total Number	e squared en of Instance	ror	25.2	971 %									
sult list (righ	nt-click for	options)		=== Detailed	Accuracy By	Class ===											
12:00:59 - fu	inctions.SM	0					Deservation	D		Mag			01				
					0.976	0.000	1.000	0.976	0.988	0.969	0.988	0.991	ckd				
				Weighted Avg	1.000 . 0.985	0.024	0.962 0.986	1.000 0.985	0.980	0.969	880.0 880.0	0.962	notckd				
				=== Confusio	n Matrix ===												
				a b <	classifie	ed as											
				0 150	b = notckd												
				-													7.
atus																	_
ок																Log	~
	S		N 4			AT/									»»	40.00	12:01 PM

Figure: 22Result of SMO classifier with Full attribute Vs Class



Volume 9 - Issue 6 - Published :

13 October 2021

W . ". ". ".	SAGAE 2000-Fold Attribute - Africa	off Wood Burehout Astronom Failed	
File Home Insert	Weka Classifier: Cost/Benefit Analysis - functions.SMO (class = ckd)		0
Cut Ca	X: Sample Size (Num) Y: True Positive Rate (Num)	X: Sample Size (Num) Y: Cost/Benefit (Num)	
Paste Format Painter	Colour: Threshold (Num) Select Instance	Colour: Threshold (Num)	
Clipboard 74	Reset Clear Open Save Jitter	Reset Clear Open Save Jitter	
L	Plot: ThresholdCurve	Plot: Cost/Benefit Curve	
	1 0.5 0 0 0 0 0.5 1		
•	% of Population % of Target (recall)	core Threshold % of Population: 0 % of Target 0 Score Threshold: 1	
	Confusion Matrix	Cost Matrix	
6	Predicted (a) Predicted (b)	Predicted (a) Predicted (b) Cost 250 Random: 250	
	0 250 co se Actual (a): ckd	0.0 1.0 Actual (a) Maximize CostBenefit	=
*	0 150 Actual (b): patrixt	1.0 0.0 Actual (b) Minimize Cost/Benefit	Ţ
	0% 37.5%	Total Population: 400 © Cost © Benefit	*
Page: 5 of 5 Words: 370			•
🚳 🙆 📳		Desktop 🎽 🔺 🌗 📜 2:39 PM 24/03/2020	

Figure: 23Weka classifier cost benefit function smo class-ckd Full attribute Vs Class



Figure: 24Weka classifier cost benefit function smo class-Not-Ckd Full attribute Vs Class

=== Classifier model (full training set) ===, SMO ,Kernel used:Linear Kernel: K(x,y) = <x,y> 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: POLYNOMICAL USING CROSS VALIDATION FOLD 4 TO 15 WITH 70% OF SPLIT CLASSIFIER Table 7:Detailed accuracy by Class with full attributes Vs Class (ckd-not-ckd)

Cross			FP		Ì	,			
Validation	Class	ТР	Rate	Precision	Recall	F-	MCC	ROC	PRC
Fold	Clubb	Rate	rute	1100151011	iteeuii	Measure	11100	Area	Area
	CKD	0.976	0.000	1.000	0.976	0.988	0.969	0.988	0.991
4	Not-	1.000	0.024	0.962	1.000	0.980	0.969	0.988	0.962
	CKD								
	Weight	0.985	0.009	0.986	0.985	0.985	0.969	0.998	0.980
	Avg								
	CKD	0.980	0.000	1.000	0.980	0.990	0.974	0.990	0.992
5	Not-	1.000	0.020	0.968	1.000	0.984	0.974	0.990	0.968
	CKD								
	Weight	0.988	0.008	0.988	0.988	0.988	0.974	0.990	0.983
	Avg								
	CKD	0.980	0.000	1.000	0.980	0.990	0.974	0.990	0.992
6	Not-	1.000	0.020	0.968	1.000	0.984	0.974	0.990	0.968
	CKD								
	Weight	0.988	0.008	0.988	0.988	0.988	0.974	0.990	0.983
	Avg								
	CKD	0.976	0.000	1.000	0.976	0.988	0.969	0.988	0.991
7	Not-	1.000	0.024	0.962	1.000	0.980	0.969	0.988	0.962
	CKD								
	Weight	0.985	0.009	0.986	0.985	0.985	0.969	0.998	0.980
	Avg								
_	CKD	0.980	0.000	1.000	0.980	0.990	0.974	0.990	0.992
8	Not-	1.000	0.020	0.968	1.000	0.984	0.974	0.990	0.968
	CKD								
	Weight	0.988	0.008	0.988	0.988	0.988	0.974	0.990	0.983
	Avg								
	CKD	0.980	0.000	1.000	0.980	0.990	0.974	0.990	0.992
9	Not-	1.000	0.020	0.968	1.000	0.984	0.974	0.990	0.968
		0.000	0.000	0.000	0.000	0.000	0.074	0.000	0.000
	Weight	0.988	0.008	0.988	0.988	0.988	0.974	0.990	0.983
	AVg	0.000	0.000	1 000	0.000	0.000	0.074	0.000	0.000
10		0.980	0.000	1.000	0.980	0.990	0.974	0.990	0.992
10		1.000	0.020	0.900	1.000	0.904	0.9/4	0.990	0.900
	Woight	0.088	0.008	0.088	0.088	0.088	0.074	0.000	0.083
	Ava	0.900	0.000	0.700	0.700	0.700	0.7/4	0.990	0.905
		0.976	0.000	1 000	0.976	ሀ ወይል	0.960	0 088	0 001
11	Not-	1 000	0.000	0.967	1 000	0.900	0.969	0.988	0.991
	CKD	1.000	0.027	0.702	1.000	0.700	0.709	0.900	0.702
	Weight	0 985	0.009	0.986	0 985	0 985	0 969	0 998	0 980
	Avø	0.705	0.007	0.200	0.205	0.705	0.707	0.770	0.200
	CKD	0.976	0.000	1.000	0.976	0.988	0.969	0.988	0.991
12	Not-	1.000	0.024	0.962	1,000	0.980	0.969	0.988	0.962
	CKD	1.000	0.021	5.762	1.000	0.200	0.707	0.200	0.702
	Weight	0.985	0.009	0.986	0.985	0.985	0.969	0.998	0.980
	Avg	0.205	0.007	0.200	0.205	0.205	0.2.02	0.220	0.200
L	'5	L	1		1		1	1	l



475

	CKD	0.976	0.000	1.000	0.976	0.988	0.969	0.988	0.991
13	Not-	1.000	0.024	0.962	1.000	0.980	0.969	0.988	0.962
	CKD								
	Weight	0.985	0.009	0.986	0.985	0.985	0.969	0.998	0.980
	Avg								
	CKD	0.980	0.000	1.000	0.980	0.990	0.974	0.990	0.992
14	Not-	1.000	0.020	0.968	1.000	0.984	0.974	0.990	0.968
	CKD								
	Weight	0.988	0.008	0.988	0.988	0.988	0.974	0.990	0.983
	Avg								
	CKD	0.980	0.000	1.000	0.980	0.990	0.974	0.990	0.992
15	Not-	1.000	0.020	0.968	1.000	0.984	0.974	0.990	0.968
	CKD								
	Weight	0.988	0.008	0.988	0.988	0.988	0.974	0.990	0.983
	Avg								





Figure: 25Detailed accuracy by class Full attribute Vs Class

Table 8:Summary of Classifier model (full training set) with full attributes 1. Test Mode: split 70% , 2. Total Number of Instances=400

Sr. No	Partkular s	4	5	6	7	8	g	10	11	12	13	14	15
ı	Time	50.0	0.03	20.0	50.0	0.02	0.02	0.03	50.0	20.0	0.02	20.0	20.0
2	build model Correctly	98.50	98.75	98.75	98.50	98.75	98.75	98.75	98.50	98.50	98.50	98.75	98.75
3	Classified Instances Incorrectly	150	1.25	1.25	1.50	1.25	1.25	1.25	150	150	150	1.25	1.25
4	Classified	0.9683	0.973	0.9735	0.968	0.9735	0.973	0.973					
5	Kappa statistik	0.015	5	0.0125	0.015	0.0125	5	5					
Б	Mean absolute error	0.1225	0.111	0.1118	0.122	0.1118	0.11 1	0.11 1					
7	Root mean squared	3.1985	2.665	2.665.6	3.198	2.6655	2.66 5	2.66 5	3.1987	3.1987	3. 1 986	2.6656	2.6657
8	error Relative absolute	25.29 7	23.09	23.09 3	25.29	23.09 2	23.0 9	23.0 9	25.29 Б	25.29 Б	25.29 Б	23.09 2	23.09 3



Figure: 26 Summary of Classifier Model with Full attribute Vs Class

===== CONFUSION MATRIX ======

The above result for all Cross Validation Folds from CVF = 4 to CVF=15 is the same Table 9:Predicted values of Confusion matrix full attributes Vs Class (ckd-not-ckd)

Sr. No.	CVF	Predicted (a)	Predicted (b)	< - Classified as
1	4	244	6	a = ckd
1	7	0	150	b = not-ckd
2	5	245	5	a = ckd



		0	150	b = not-ckd
3	6	245	5	a = ckd
5	Ŭ	0	150	b = not-ckd
4	7	244	6	a = ckd
1	,	0	150	b = not-ckd
5	8	245	5	a = ckd
	Ŭ	0	150	b = not-ckd
6	9	245	5	a = ckd
0	,	0	150	b = not-ckd
7	10	245	5	a = ckd
,	10	0	150	b = not-ckd
8	11	244	6	a = ckd
0		0	150	b = not-ckd
9	12	244	6	a = ckd
	12	0	150	b = not-ckd
10	13	244	6	a = ckd
10	10	0	150	b = not-ckd
11	14	245	5	a = ckd
**		0	150	b = not-ckd
12	15	245	5	a = ckd
12	15	0	150	b = not-ckd



Unsupervised – Clustering – Hierarchical –Euclidean & Manhattan Distance Function using No of Clusters 2,3,4& 5 test results: Full Attributes



Step 1: Using Euclidean Distance function No of Clusters 2,3,4& 5

Open new Open Out Open Out Open Out Open Out Out Date Out Ou		aerate II	ado) [Save Save
ren relation: Relation: Chronic, Ködney, Disease_(RS Walase)-weika. filters unsupervised attribute. Nu				Jave
Choose Numeric Totominal - R fristiat Apply rent relation				
rer teletion Selected attribute Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.unsupervised attribute. YS Attributes: 25 Instances: 400 Distinct. 7 Unique: 1(0%) Instances: 400 Distinct. 7 Distinct. 7 Instance: 400 Distinct. 7 Distinct. 7 Instinte: 400	Choose NumericToNominal -R first-last			Apply Stor
Relation: Chronic_Kidney_Disease_[RSWaise}-weka.filters unsupensised attribute.Nu Attributes: 25 Instances: 400 Distinct: 7 Unque: 1 (0%) Instance: 400 Distinct: 7 Unque: 1 (0%) Instance: 400 Distinct: 7 Unque: 1 (0%) Instance: 400 Distinct: 7 Unque: 10(%) Instance: 400 Instance: 400 Instance: 400 Instance: 400 Instance: 400 Instance: 400 Instance: 400	rrent relation	Selected attribute		
No. Label Count Weight 1 0 199 199.0 2 1 44 44.0 3 1.016949 46 46.0 4 3 1.016949 46 43.0 5 3 43 43.0 5 5 5 5 5 6 7 5 1 1.0 7 5 1 1.0 8 5 5 5 5 6 7 5 1 1.0 7 5 1 1.0 1.0 9 5 5 5 5 5 10 5 7 5 1 1.0 7 5 1 1.0 1.0 1.0 110 5 5 5 5 5 10 5 7 5 1.0 1.0 111 5 5 5 5 1.0 12 5 5 1.0 1.0 1.0 12 5 5 1.0 1.0 1.0 13 5 5 1.0 1.0 1.0 1	Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.unsupervised.attribute.Nu Attributes: 25 Instances: 400 Sum of weights: 400	Name: al Missing: 0 (0%)	Distinct 7	Type: Nominal Unique: 1 (0%)
All None Invert Patern 1 0 199 1990 2 1 44 440 3 1016949 46 46.0 4 2 43 43.0 5 3 43 43.0 6 4 24 24.0 7 5 1 10 6 10 5 1 9 0 0 1 10 5 1 10 11 0 1 10 12 5 1 10 13 50 1 10 10 5 1 10 9 0 0 1 11 5 1 10 12 5 1 10 13 50 1 10 14 od 1 1 15 perov 1 1 16 pcv 1 1 15 pcv 1 1 15 pcv 1 1 16 pcv 1 1	ributes	No. Label	Count	Weight
All None Invert Pattern 0. Name 44 44.0 3. 30 45 46.0 4. 2. 43 43.0 2. bp 3. 43.0 3. 5. 3. 43 43.0 5. 9.0 7.5 1 1.0 6. bc 7.5 1 1.0 7. pc 9.0 0.0 1.0 8. pcc 9.0 0.0 1.0 9. bar 0.0 1.0 1.0 10. bgr 1.0 1.0 1.0 11. bu 1.0 1.0 1.0 12. sc 1.0 1.0 1.0 13. sod 1.0 1.0 1.0 10. bgr 1.0 1.0 1.0 12. sc 1.0 1.0 1.0 13. sod 1.0 1.0 1.0 13. sod 1.0 1.0 1.0 13. sod 1.0 1.0 1.0 14. od 1.0 1.0 1.0 15. prov		1 0	199	199.0
All None Invert Pattern 3 1.016949 46 46.0 No. Name 43 43.0 1 age 5 3 43 43.0 2 bp 5 3 43 43.0 3 bp 5 3 43 43.0 5 3 43 43.0 6 6 fbc 7 5 1 1.0 6 fbc 7 5 1 1.0 7 pc 6 fbc 7 5 9 ba 0 0 1 1.0 10 pgr 13 sod 1 1 11 put 1 1 1 1 15 pero 1 1 1 1 15 pcv 1 1 1 1 15 pcv 1 1 1 1 16 pcv 1 1 1 1 16 pcv 1 1 1 1 16 pcv 1 1 1 1 17 1 1 1 <		2 1	44	44.0
Name 4 2 43 430 1 age 430 430 2 bp 6 4 24 2 bp 1 10 3 sg 1 10 4 at 10 10 5 bc Class: class (Nom) V (su 7 pc 10 bgr 10 bgr 11 10 12 sc 13 sod 14 pd at at 15 hermove at at 16 pcv at at 13 sod at at 15 hermove at at 16 pcv at at 16 pcv at at	All None Invert Pattern	3 1.016949	46	46.0
100 Name 10 Name 11 986 2 bp 3 53 4 al 5 3 6 4 7 5 10 bp 3 sg 4 al 5 3 6 7 5 3 7 pc 9 ba 10 bgr 11 pc 9 ba 10 bgr 11 pu 12 scd 13 sod 14 pot 15 hermove Remove K		4 2	43	43.0
1 gge 2 bp 3 sg 4 ad 5 su 6 r 6 r 7 p 8 pcc 9 pa 10 bgr 11 bu 12 sc 13 sod 14 pot 15 hemo 16 pov	io. Name	5 3	43	43.0
2 bp 3 59 4 31 5 su 6 rtc 7 5 1 10 7 5 1 10 (class: class (Nom) Visu 7 pc 8 pcc 9 ba 9 ba 10 bgr 11 bu 12 sc 13 sod 14 pot 15 hemo 16 pcv Remove Remove Log	1 age	6 4	24	24.0
3 9 4 0 5 9 6 rbc 7 pc 8 pcc 9 ba 10 gr 11 bu 12 sc 13 sod 13 sod 14 pot 15 hemo Remove Remove K	2 bp	7 5	1	1.0
5 su 6 rbc 7 pc 9 ba 9 ba 10 gyr 11 bu 12 sc 13 sod 15 hemo 16 pcv Remove Remove k	3 sg 4 al			
0 000 7 0 8 pcc 9 ba 10 bgr 11 bu 12 sc 13 sod 14 pot 15 hemove Remove tus K	5 su	Class: class (Nom)		Visualiz
10 bgr 10 bgr 11 bu 12 5c 13 sod 14 pot 15 hemo 16 pov Remove us K	7 00	Class. class (Nonit)		- Haddin
9 bc 9 bc 10 bgr 11 bu 12 sc 13 sod 14 pot 15 hemo 16 pcv Remove K	8 pc			
10 bgr 11 bu 12 sc 13 sod 14 pot 15 hemo 16 pov Remove us K		199		
11 5 5 C 13 5 sod 14 pol 15 hemo 16 pov Remove K K	10 brr			
12 sc 13 sod 14 pot 15 hemo 16 pov Remove Remove k	11 bu			
13 od 14 pot 15 hemo 16 pov Remove K K	12 sc			
14 pot 15 hemo Remove Remove K Log	13 Sod			
15 hemo 18 pov Remove tus K	14 pot			
16 pcv Remove 44 45 43 49 44 45 44 45 44 45 45 44 45 45 45 44 45 <td>15 hemo</td> <td></td> <td></td> <td></td>	15 hemo			
Remove 41 43 43 44 14 14 14 14 14 14 14 14 14 14 14 14	16 🗌 pcv 💽			
tus К	Remove	44	40 43	43
htus DK				24
ж Log	tus			
	ж			Log

Figure: 28Preprocess Full attribute Vs Class

Tuester mode	Clusterer oute	😋 weka.qui.GenericObjectE	Editor	
Use training set Supplied test set SeL.	Cluster 0 ((((((0.0	weka.clusterers.Hierarchic	alClusterer	10,0.010)
Classes to clusters evaluation (Nom) class ✓ Store clusters for visualization	Cluster 1 (((((0.0:)	Hierarchical clustering	g class. More Capabilities	
Ignore attributes Start Stop	Time takes === Model Cluster 0 (0.0:0,0.1	debug distanceFunction	False Choose EuclideanDistance -R first-last False	
lesuit list (right-click for options)	Cluster 1 ((((((0.0	doNotCheckCapabilities linkType	False Single))); :1, ((0.0:1, (((((0.0:0, 0.0:0):0, 0.0:0):0, 0.0:0):1, (((((
	Time take: Clustered	numClusters printNe <mark>Sets t</mark>	the number of clusters	
	1 9	Open	Save OK Cancel	,

Figure: 29Hierarchical clustering class with Euclidean function of Full attribute Vs Class with no of cluster 2,3,4& 5



Weka Explorer	
Clusterer	
Choose HierarchicalClusterer -N 2 -L SINGLE -F	-A "weka.core.EuclideanDistance -R first-last"
Cluster mode	Clusterer output
Use training set Supplied test set Set Percentage split % 70 Classes to clusters evaluation (Nom) class Store clusters for visualization	Cluster 0 ((((((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0):0,(0.0:0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,(0.0:0,0.0:0):0):0,(0.0:0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0):0,(0.0:0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,(0.0:0,0.0:0):0):0,0.0:0):0):0,(0.0:0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0):0,(0.0:0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0):0,(0.0:0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0):0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0,0.0:0):0):0,(0.0:0,0.0:0):0):0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0);0,0.0:0) Cluster 1 (((((0.0:0,0.0:0):0,0.0:0):0,((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0);0,0.0:0);0,0.0:0);0,0.0:0);0,0.0:0;0,0,0;0,0,0;0,0,0;0,0,0;0,0,0;0,0;0,
Ignore attributes Start Stop Aesult list (right-click for options)	dm cad appet pe ane ((((((0.0:0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0):0,0.0:0):1,((((((Select Pattern Cancel
	Time taken to build model (percentage split) : 0.09 seconds Clustered Instances 0 21 (18%) 1 99 (83%)
L	
OK	

Figure: 30No of 02 cluster of hierarchical clustering class clustering output no of cluster 2,3,4& 5 STEP 1:

(Euclidean Distance) = 2 Cluster = Full Attributes

=== Run information ===

weka.clusterers.HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.EuclideanDistance -R Scheme: first-last" Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last Instances: 400 Attributes: 25 age bp sg al su rbc pc pcc ba bgr bu sc sod pot hemo pcv wbcc rbcc htn dm cad appet pe ane Ignored: class Test mode: Classes to clusters evaluation on training data === Clustering model (full training set) === Cluster 0 Time taken to build model (full training data) : 0.73 seconds === Model and evaluation on training set === Clustered Instances 0 399 (100%) 1 1 (0%)

Class attribute: class Classes to Clusters: 0 1 <-- assigned to cluster



249 1 | ckd 150 0 | notckd Cluster 0 <-- ckd Cluster 1 <-- No class Incorrectly clustered instances : 151.0 37.75 % STEP 2: (Euclidean Distance) = 3 Cluster = Full Attributes Time taken to build model (full training data) : 0.73 seconds === Model and evaluation on training set === Clustered Instances 0 398 (100%) 1 1 (0%)

2 1 (0%)

Class attribute: class Classes to Clusters:

0 1 2 <-- assigned to cluster 248 1 1 | ckd 150 0 0 | notckd

Cluster 0 <-- ckd Cluster 1 <-- No class Cluster 2 <-- No class

Incorrectly clustered instances : 152.0 38 % STEP 3:

(Euclidean Distance) = 4 Cluster = Full Attributes

Time taken to build model (full training data) : 0.7 seconds

=== Model and evaluation on training set === Clustered Instances

0 397 (99%) 1 1 (0%) 2 1 (0%)

3 1 (0%)

Class attribute: class Classes to Clusters:

0 1 2 3 <--- assigned to cluster 247 1 1 1 | ckd 150 0 0 0 | notckd

Cluster 0 <-- ckd Cluster 1 <-- No class

Cluster 2 <-- No class Cluster 3 <-- No class

Incorrectly clustered instances : 153.0 38.25 % STEP 4:



(Euclidean Distance) = 5 Cluster = Full Attributes

Time taken to build model (full training data) : 0.69 seconds

=== Model and evaluation on training set === Clustered Instances

0 396 (99%)

1 1 (0%)

2 1 (0%)

3 1 (0%)

4 1 (0%)

Class attribute: class Classes to Clusters:

0 1 2 3 4 <-- assigned to cluster 246 1 1 1 1 | ckd 150 0 0 0 0 | notckd

Cluster 0 <-- ckd Cluster 1 <-- No class Cluster 2 <-- No class Cluster 3 <-- No class Cluster 4 <-- No class

Incorrectly clustered instances : 154.0 38.5 %



Figure: 31Hierarchical clustering with Manhattan distance Function of Dendrogam

Figure 3: Performance accuracy by Class and Confusion matrix by decision tree. Step 2: Using Manhattan Distance function No of Clusters 2,3,4& 5



Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize				
Open file	Open DB Gener	rate U	ndo Ec	it Save
filter				
Choose NumericToNominal -R first-last				Apply Stop
Current relation		Selected attribute		
Relation: Chronic_Kidney_Disease_(RS Walse)-weka.filters.unsupervised.at Instances: 400	tribute.Nu Attributes: 25 Sum of weights: 400	Name: al Missing: 0 (0%)	Distinct 7	Type: Nominal Unique: 1 (0%)
Attributes		No. Label	Count	Weight
		1 0	199	199.0
All None Invert	Pattern	2 1	44	44.0
	ratem	3 1.016949	46	46.0
and areas	11	4 2	43	43.0
No. Name		0 3	43	43.0
1 🧾 age	A	7 6	24	24.0
2 📃 bp		/ 5	1	1.0
3 🛃 sg				
4 al				
5 🛄 su		(m.).		
6 📃 rbc		Class: class (Nom)		Visualize A
7 🛄 pc				
8 🛄 pcc				
9 🛄 ba		199		
10 bgr				
11 🔟 bu				
12 🔄 sc				
13 🛄 sod				
14 🔄 pot				
15 🔜 hemo				
16 pcv	T			
Pemaie	1	44	46 43	43
Kentove				24
tatus				
ок				Log 🛷 X
8 🧟 🔚 🔽 🧿 📼 🐼				Desktop ဳ 🔺 🌓 📷 11:50 PM

Figure: 32Preprocess data set with full attributes for Manhattan Distance function

Weka Explorer		Part along him	Name and Address of the Owner, where	Statement of the local division in which the local division in the local division in the local division in the		
Preprocess Classify Cluster Associate Select	attributes Visuali	ze				
Clusterer						
Choose HierarchicalClusterer -N 2 -L SINGLE -P	-A "weka.core.Manh	attanDistance -R first-last"				
Cluster mode	Clusterer outp	😡 weka.gui.GenericObject	Editor		×	
Use training set Supplied test set Set. Percentage split % 70 Classes to clusters evaluation (Nom) class	Model Clustered 0 39 1 2	About Hierarchical clusterin	g class.	More Capabilities	}	ŕ
Ignore attributes Start Stop Result list (right-click for options)	Class att Classes t 0 1 246 1 150 0 Cluster 0 Cluster 1 Cluster 2 Cluster 4 Incorrect	debug distanceFunction distanceIsBranchLength doNotCheckCapabilities linkType numClusters printNewick Open	False Choose ManhattanDistance - f False False SINGLE Sets the number of cluster True Save OK	R first-last		
Status OK						Log 🛷 x0
🚳 🖉 📜 💽 🧕		W				Desktop 🎽 🔹 🌓 📜 12:14 AM 25/03/2020

Figure: 33Hierarchical Clustering class with Manhattan Distance function for no of cluster 2,3,4& 5



Figure: 34Result of Hierarchical Clustering class with Manhattan Distance function for no of cluster 2,3,4& 5

STEP 1:

```
(Manhattan Distance) = 2 Cluster = Full Attributes
=== Run information ==
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
Instances: 400
Attributes: 25
Time taken to build model (full training data) : 0.72 seconds
=== Model and evaluation on training set === Clustered Instances
0
       399 (100%)
       1 (0%)
1
Class attribute: class Classes to Clusters:
0 1 <-- assigned to cluster
249 1 | ckd
150 0 | notckd
Cluster 0 <-- ckd Cluster 1 <-- No class
Incorrectly clustered instances :
                                       151.0 37.75 %
STEP 2:
(Manhattan Distance) = 3 Cluster = Full Attributes
Time taken to build model (full training data) : 0.71 seconds
=== Model and evaluation on training set === Clustered Instances
```

- 0 398 (100%)
- 1 1 (0%)



2 1 (0%) Class attribute: class Classes to Clusters: 0 1 2 <-- assigned to cluster 248 1 1 | ckd 150 0 0 | notckd Cluster 0 <-- ckd Cluster 1 <-- No class Cluster 2 <-- No class Incorrectly clustered instances : 152.0 38 % STEP 3: (Manhattan Distance) = 4 Cluster = Full Attributes Time taken to build model (full training data) : 0.73 seconds === Model and evaluation on training set === Clustered Instances 397 (99%) 0 1 (0%) 1 2 1 (0%) 3 1 (0%) Class attribute: class Classes to Clusters: 0 1 2 3 <-- assigned to cluster 247 1 1 1 | ckd 150 0 0 0 | notckd Cluster 0 <-- ckd Cluster 1 <-- No class Cluster 2 <-- No class Cluster 3 <-- No class STEP 4: (Manhattan Distance) = 5 Cluster = Full Attributes === Model and evaluation on training set === Clustered Instances 0 396 (99%) 1 1 (0%) 2 1 (0%) 3 1 (0%) 4 1 (0%) Class attribute: class Classes to Clusters: 012 3 4 <-- assigned to cluster 2461111 | ckd 150 0 0 0 0 | notckd Cluster 0 <-- ckd Cluster 1 <-- No class Cluster 2 <-- No class Cluster 3 <-- No class Cluster 4 <-- No class

Incorrectly clustered instances : 154.0 38.5 %





Volume 9 - Issue 6 - Published : 🗰 13 October 2021

Figure: 35Result of Hierarchical Clustering class with Dendrograph Figure 35 What shows?

IV. RESULTS AND DISCUSSION

{a}Comparative statement of Supervised SMO-SVM- Classifier with one, two & full attributes with 70% split using Calibrator: Logistic and Kernel= Polynomial function

The research, the kidney dataset processed with different attributes (25), which contains 400 rows, i.e., instances and 25 attributes, means columns. The researcher has select every attributes to displays type of attributes, the type means Nominal, how many missing values present in the data set for each attribute viz. instances, how many distinct values are present in the dataset, distinct means different values, if we select attribute is shown Nom- in front of attribute- nom means nominal type. The numeric data gives summary of the overall data in the form of descriptive statistics. Also, if data is qualitative then it treated as an attribute class and it shows in the form of label count and its weight in the form of true/false or yes/no.

Table 10: Correctness by Class values



Table 10: Correctness by Class values										
	No of	Class	TP	FP	Precisi	Reca	F-	MCC	ROC Area	PRC
	Attribut		Rate	Rate	on	11	Measure			Area
	es									
	01	CKD	0.588	0.000	1.000	0.58	0.741	0.59		0.845
						8		0	0.794	
*CVF=4	(htnVs	Not-CKD	1.000	0.412	0.593	1.00	0.744	0.59	0.794	0.593
						0		0		
	CIASS)	Weight	0.743	0.155	0.847	0.74	0.742	0.59	0.794	0.751
	Ανσ					3		0		
	5					-		-		
	02	CKD	0.737	0.00	1.000	0.73	0.84	0.712		0.904
"OVE				0		7	8		0.868	
CVF										
	appetV	Weight	0 83	0 00	0 886	0 83	0 83	0 172	0 868	0 824
	Class)	Arra	0.00	۵.03 د	0.000	0.03	6.05	0.1/2	0.000	0.024
		er var	3	0		3	0			
	24	CKD	0.98	0.00	1.000	0.98	0.99	0.974		0.992
			0	0		0	0		0.990	

* Found max and same accuracy for 5 to 15 CVF

** Found max and same accuracy for 5 to 15 CVF

*** 5,6,8,9,10,14,15 found max and same accuracy CVF

Table 10shows the correctness by Class valuesSupervised- classify - SMO- SVM- with attributes 2,3 & Full Attributes – Applying Cross validation foldfrom 4 to 15 with 70% split, and applying the Calibrator: Logistic and Kernel: Polynomial function and it shows the comparative study of the result and found the accurate prediction value, therefore research perfectly state that the increasing no of cross validation folds with respective increasing attributes the accuracy of the correctness of the class values increases shows in the form of ROC Area CVF=4 ROC Area = 0.794 using one attribute, CVF=4 ROC Area = 0.868 using two attributes and CVF=5 ROC Area=0.990. Hence we found best and Novel Predictive Module with the help of SMO- SVM-Classifier Class Module on the basis of Calibrator and Kernel function.



Figure: 36 Comparative study of Correctness by class values of SMO-SVM- Classifier with 1,2 and full attributes with 70% and CVF from 4 to 15.

Table 11:Summary	of Classifier model	(Train set data)
------------------	---------------------	------------------

		01	02	24
		(htnVs Class)	(htn, appetVs Class)	(Full Vs Class)
Sr.	Particulars	CV Folds	CV Folds	CV Folds
No.		4 Found max and	4 Found max and	5,6,8,9,10,14,15
		same	same	found
		accuracy for 5 to 15	accuracy for 5 to 15	max and same
		CVF	CVF	
				accuracy CVF
1	Test mode: 70% train, 30%	70.0% train	70.0% train	70.0% train
	test			
2	Correctly Classified	74.25	83.33	98.75
	Instances			
3	Incorrectly Classified	25.75	16.66	1.25
	Instances			
4	Kappa statistic	0.517	0.6725	0.9735
5	Mean absolute error	0.2575	0.1667	0.0125
6	Root mean squared error	0.5074	0.4082	0.1118
7	Relative absolute error	54.90	35.6241	2.6656
8	Root relative squared error	104.81	84.6877	23.094
9	Total Number of Instances	400	400	400

(CVF= Cross Validation Folds)

Table 11 shows the Summary of classifier model with train set data, of SMO- classifier by applying one, two and full 25 attributes to test the result and getting some of the useful and accurate result, so as to use our research to further researcher for further prediction of data and effectively useful of the SMO – SVM classifier algorithm, further table 11 shows the how accuracy will be increased on the basis of different parameters like Cross Validation Folds just testing from 4 to 15, accuracy of the result is higher as compared to the earlier result in the form of Correctly Classifier Instances is 74.25 % , 83.33% and 98.75% respectively. Therefore also found the best Novel Summary of Classified module with highest accuracy of CCI.





Figure: 37Comparative study of Summary of Classifier Module of SMO-SVM- Classifier with 1,2 and full attributes with 70% and CVF from 4 to 15.

Table 12:Comparative statement of Confusion Matrix SMO-SVM- Classifier with one, two & full attributes with 70% split using Calibrator: Logistic and Kernel= Polynomial function

Sr. No.	No of	CVF	Predicted (a)	Predicted (b)	< - Classified as
	Attributes				
1	One Attribute	4-15	147	103	a = ckd
	htnvs Class		0	150	b = not-ckd
		4	147	103	a = ckd
2	Two Attributes		0	150	b = not-ckd
	Htn, appetVs	5-15	172	78	a = ckd
	Class		0	150	b = not-ckd
		4,7,11,12,13	244	6	a = ckd
	Full Attributes		0	150	b = not-ckd
3	24 Vs Class	5,6,8,9,10,14,15	245	5	a = ckd
			0	150	b = not-ckd

(CVF = Cross Validation Folds)

Table 12 shows the Comparative statement of Confusion Matrix SMO-SVM- Classifier with one, two & full attributes with 70% split using Calibrator: Logistic and Kernel= Polynomial function, it plot the threshold Curve and Cost benefit curve for class ckd and Not-ckd in the form of % population of confusion matrix and cost matrix. Confusion matrix of predicted a & b also cost matrix a & b with the random and gain value. The researcher also find the research output is using full length of attributes it creates minimum Confusion matrix with accurate predicted class value with minimum no of iterations of clusters.

{b}Comparative statement of Un-Supervised Hierarchical Clustering algorithm using Euclidean and Manhattan distance function for the no of cluster 2,3,4 & 5.

Comparative statement of the Hierarchical clustering algorithm using Euclidean nad Manhattan distance function result with accuracy.



Euclidean Distance:

=== Run information ===

Scheme: weka.clusterers.HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.EuclideanDistance -R first-last" Relation:Chronic_Kidney_Disease_(RS Walse)-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

I Manhattan Distance:

=== Run information ===

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 f	ull attributes	with 70% split. (Cluster 2,3,48	τ 5)					

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

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.

V. 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 algorithmNovel discovered module is the best truthful classifier with 98.75% accuracyi.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 measuredRBC 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, thee 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.

VI. 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.

VII.REFERENCES

- 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 (ICCCT), 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. SaiSabitha, and AbhayBansal. "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, ZengYuqun, 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 ThanhAnhLuong, Dept of the computer. Sci&Engi, uni, at Buffalo Buffalo, NY, USAA K- Means Approach to Clustering disease Progressions IEEE Keywords, Sep 2017
- [6]. Güllüoğlu, SabriSerkan. "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, SabitaKLhatri, 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, NakornthipPrompoon, 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.
- [14] . https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

