International Journal of Engineering Research and Technology.
ISSN 0974-3154 Volume 11, Number 11 (2018), pp. 1781-1794

http://www.irphouse.com

Chronic Kidney Disease: A Predictive model using Decision Tree

Vikas Chaurasia[1], **Saurabh Pal**[1,*], **B.B. Tiwari**[2]
[1]Dept of MCA,VBS Purvanchal University, Jaunpur, UP, India
[2]Dept. of Electronics, VBS Purvanchal University, Jaunpur, UP, India
**Corresponding author*

Abstract

Data mining may be utilized in healthcare industry in actual fact to “mined” clinical information to get hidden information for intelligent and effective decision making. Advanced data processing techniques in discovery of hidden patterns and relationships may be a fruitful as remedy to the present state of affairs, principally deals with Prediction of Chronic urinary organ illness. Information covers several attributes like blood, urine, cardiovascular disease check, and external symptoms applied to predict chronic urinary organ illness.

For getting higher result several parameters are accustomed interaction between measured parameters and that is get from data processing approach, processing and information transformation of the survival of the patient. Decision Tree algorithms are used for extracting information within the kind of a collection of decision rules. The decision-making algorithm is employed to predict the survival of the CKD patient and additionally who is new and unseen. For the medical purpose more important parameters are known. During this analysis paper concept is introduced check and apply the information assortment from UCI Machine Learning Repository Chronic_Kidney_Disease information Set_files. The procedure results are evaluated during this research paper with medical significance.

Keywords: Chronic kidney disease, data mining, Clinical information, data Transformations, Decision-making algorithm

1. INTRODUCTION

Worldwide out of total population, 10% are affected by chronic kidney disease (CKD) and since of they don’t have reasonable medical access thus millions die annually because of economic condition [1]. In China and India, it's calculable that range of patients of kidney failure can increase in developing countries, wherever the amount of unbalanced older folks are increasing extremely [2]. In several low and middle-income countries, because of elevated charge of healing with dialysis or urinary organ transplantation creates a wealth load for the bulk of people. In alternative countries, several majority people cannot afford the treatment at all, thus within the result the death of over one million people from untreated kidney failure annually [3]. Chronic urinary organ illness is turning into a growing problem in each developed and developing world. The estimated range of deaths due to chronic diseases in India will increase from 3.78 million (40.4% of all deaths) in 1990 to 7.63 million (66.7% of all deaths) in 2020[4]. Chronic kidney disease, or chronic kidney failure, describes the step by step loss of urinary organ operate. The defecation in piss may be a process of filtration of excess and wastes fluids to Kidneys from blood. In highly developed stage of CKD, risky levels of wastes, electrolytes, and fluid will construct in body. There could also be little indication or sign within the early step of chronic kidney disease. Once urinary limb operate is considerably impair chronic kidney disease may not become clear. Healing of CKD solely focuses on retardation the method of the urinary organ harm. CKD will propagate to end-stage of kidney failure that is life threatening while not artificial filtering (dialysis) or a kidney transplant. Commonest symptoms of nephropathy could include:

- Sleep troubles
- Vomiting
- Loss of appetite
- Tiredness and weakness
- Nausea
- Changes in how much urinate
- Pain in chest by the builds from fluid round the lining of the center
- Decreased mental sharpness
- Twitches and cramps in muscle
- Swelling of feet and ankles
- Hypertension that is uncontrolled
- Persistent cutaneous sensation
- Breath shortness if lungs stuffed up with fluid

Impairing of urinary organ operate produce chronic kidney disease that’s why urinary organ harm to worst condition inside some months or years.
There are some conditions that will reason for chronic kidney disease:

- High blood pressure
- Type one or sort a pair of diabetes
- An irritation of the kidney's filtering units
- Interstitial nephrosis, an inflammation within the encompassing structures and kidney's tubules
- Delayed obstruction of the region, from circumstances like enlarged prostate, urinary organ stones and a few cancers
- Recurrent urinary organ infection
- Polycystic kidney disease
- A condition that causes piss to back up into kidneys

Circumstance, that enhance risk of CKD include:

- High blood pressure
- Older age

- Heart and blood vessel disease
- Smoking
- Obesity
- Abnormal urinary organ structure
- Diabetes
- For the Native Asian-American or American, African-American
- History of family of kidney disease

CKD will have an effect on the majority a part of body. Potential complications could include:

- Decreased dysfunction, erectile, sex drive or reduced fertility
- A boost in potassium levels in blood that might impair heart's ability to control and should be critical
- Risk of bone fractures for weak bones
- Anemia
- Fluid retention that could lead on to high blood pressure, swelling in arms and legs or fluid in lungs (pulmonary edema)
- Irreversible harm of kidneys, ultimately requiring either kidney transplant or dialysis a for survival
- Damage central systema nervosum, which can modification in temperament or issue in concentrating
- Heart and vessel (cardiovascular) illness
- Decreased immune reaction, that makes additional prone to infection
- Pregnancy complications that carry risks for developing craniate and mother

To reduce these risk for developing urinary organ disease:

- Manage medical conditions with doctor's facilitate. If anyone has illness or conditions to extend the chance of kidney disease, check with doctor to regulate them.
- Follow directions on medications. Once using nonprescription pain relievers, like non steroidal anti-inflammatory drug, acetylsalicylic acid and acetaminophen, follow the directions on the wrap up. One ought to avoid taking too several pain relievers that will result in urinary organ harm.
- Don't smoke. If you are a smoker avoid smoking, cigarette smoking could harm urinary organs and build existing kidney harm worse.
- Maintain a healthy weight. For the healthy weight one ought to work to take care of it by being physically active entire days of the week. It additionally has to slenderize and speak with doctor regarding ways that maintain the healthy weight loss. This involves increasing reducing calories and daily physical activity.

With early diagnosis and treatment, chronic kidney disease may be treated, it's potential to cut down or stop the progression of nephropathy.

Survival analysis of CKD using normal applied math tool, like tree-based strategies for regression and classification “population-based” models, these keep segmenting or stratifying the predictor space into variety of straight forward regions. The predictor space involves stratifying or segmenting into variety of straightforward regions. It generally uses the mean or the mode of the training observations so as to form a prediction for a given observation, within the region to that it belongs. The predictor space may be summarized during a tree, by the set of splitting rules accustomed segment; these kinds of approaches square measure referred to as call tree strategies [5, 6]. Searching for hidden patterns in information can represent valuable information, which can result in medical investigations, for example, certain limitations of parameter values, which result in an extended existence time. For this purpose, a knowledge mining approach was used to establish relevant factors that predict a person's individual living by the time of intermediary existence.

2. RELATED WORKS

In present scenario study and research goes within the field of medical diagnosis. Variety of work has been done on diseases like diabetes, Cancer and coronary failure using many data processing techniques. a number of the recent analysis supported alternative unwellness and chronic kidney disease prediction using varied techniques of information mining is listed below;

Ani R et al., (Ani R et al.2016) planned a approach for prediction of CKD with a changed dataset with 5 environmental factors. The methodology introduced during this paper for the prediction is perfect Fuzzy-K nearest Neighbor. This research work show because the higher classifier with two totally different categories of diseases with excellent classification rate. they need taken three classifier, support vector machines, artificial neural network and K nearest neighbors for comparative study between them and also the metrics that's used for the analysis are accuracy, precision, recall, specificity and F-Measure. so it may be say that the analysis for identification of CKD, the planned methodology is well organized to get the right classifier. The conclusion half shows that the analysis leading to high classification accuracy rate for the OF-KNN machine learning tool.

Zeinab Sedighi et al., (Sedighi Z. et al. 2015) during this analysis paper, they planned the result of feature choice in CKD classification. They used some special classification approach with the assistance of filter and wrapper based mostly feature choice techniques and compared in terms of classification accuracy. By choosing a number of the options in some applications decrease quality and time period of the classification model that resound or have low correlation with alternative characteristics. so the investigator shows that the genetic algorithmic program compares some common feature choice strategies using associate degree ensemble classification and also the accuracy of classification is nice as compared with the initial options.

Murat Koklu et al., (Koklu K. et al. 2016) planned a classification methodology of CKD with the assistance of Naive Thomas Bayes, SVM, C4.5 and Multilayer Perceptron. in several classification methodology they show that Naive bayes is that the best performer among the classifier they used and criteria for choosing the most effective classifier are accuracy, performance analysis measures (TP FP Tennessee

FN), precision, recall:, F-measure, ROC curve, confusion matrix.
Vikas Chaurasia et al., (Chaurasia V. et al. 2018), using data processing techniques, planned a test paper on the prediction of benign and malignant breast cancer, which gave the three well-liked data processing techniques useing, RBF Network, took J48 and took a look to indicate that Nive Bayes' accuracy is ninety seven and alternative algorithmic program is a smaller amount entertainer just in case of breast cancer survivals. He used a binary clear existence variable, which was calculated from the variable within the raw dataset, where it is depicted with the value of "1" to represent the fatality and '0' for curable. Therefore, to achieve fair prediction accuracy of three strategies, 10 times cross verification process was used, which divided the dataset into 10 mutually exclusive divisions (K-folds) that employ a sample technique. It supplied with less favorable prediction performance measures to match the three models. Sensitivity results show that the forecast issues 'class' is the most important predictor, far and away [12, 13].
Saurabh Pal et al., (Pal S. et al. 2017) given a paper on the performance of the students who consume alcohol throughout gaining their higher education. They used totally different data processing algorithms BFTree, J48, RepTree and simple Cart to match the result to spot that student who really want associate degree special attention throughout their studies. correct and short prediction of attributes that facilitate the education professionals in addition as establishments to predict those students who have the worst performance; they'll develop the special techniques to enhance them with special attention confidently. The most effective algorithmic program supported the student’s knowledge is BFTree Classification with accuracy of eightieth and also the total time taken to make the model is at 0.83 seconds.

Andrew Kusiak et al., (Kusiak A. et al. 2005) given an information mining approach for the patients whose excretory organ dialysis goes on and through this era what's their survival time. For this purpose, they show that data processing, knowledge change, knowledge division, and decision-making algorithms are helpful for live predictions of dialysis patients. For anyone patients the classification accuracy raised enough to 75 – 85%. When transformation of information, the accuracy of the classification is raised by 11%. For the take a look at knowledge set call tree algorithmic program turn out 67% correct prediction for the survival time and additionally RS algorithmic program turn out 56% correct prediction of the time for the easy in addition as weighted voting scheme on the take a look at knowledge set.

3. DATA COLLECTION AND PREPROCESSING

Data collected from UCI Machine Learning Repository Chronic_Kidney_Disease Data Set_files. This data set has 400 instances which have total number attributes 24 and 1 class, 11 numeric and 14 nominal. The relevant information about attributes is given in the following Table 1.

Table1. Attribute Information

Attribute symbols	Attribute description	Attribute type
age	age in years	numerical
bp	Blood Pressure (bp in mm/Hg)	numerical
sg	Specific Gravity (1.005,1.010,1.015,1.020,1.025)	nominal
al	Albumin (0,1,2,3,4,5)	nominal
su	Sugar (0,1,2,3,4,5)	nominal
rbc	Red Blood Cells (normal, abnormal)	nominal
pc	Pus Cell (normal, abnormal)	nominal
pcc	Pus Cell clumps (present, not present)	nominal
ba	Bacteria (present, not present)	nominal
bgr	Blood Glucose Random (bgr in mgs/dl)	numerical
bu	Blood Urea (bu in mgs/dl)	numerical
sc	Serum Creatinine (sc in mgs/dl)	numerical
sod	Sodium (sod in mEq/L)	numerical
pot	Potassium (pot in mEq/L)	numerical
hemo	Hemoglobin (hemo in gms)	numerical
pcv	Packed Cell Volume	numerical
wc	White Blood Cell Count (wc in cells/cumm)	numerical
rc	Red Blood Cell Count (rc in millions/cmm)	numerical
htn	Hypertension (yes, no)	nominal
dm	Diabetes Mellitus (yes, no)	nominal
cad	Coronary Artery Disease (yes, no)	nominal
appet	Appetite (good, poor)	nominal
pe	Pedal Edema (yes, no)	nominal
ane	Anemia (yes, no)	nominal
class	Class (ckd, notckd)	nominal

4. METHODOLOGY AND EXPERIMENTAL RESULTS DISCUSSION

In the following sections it is mentioned the methodology of the research which dealt with the using method i.e. decision tree.

4.1 decision trees

Decision tree is a supervised algorithm for learning in classification problems, which has a pre-defined target variable which is mostly used. Decision tree works for both specific and continuous input and output variables. In this method, Decision tree can be applied to both classification and regression problems, which divides the population or sample into two or more homogeneous sets called sub-population based on the most important splitter in the input variable. [7].

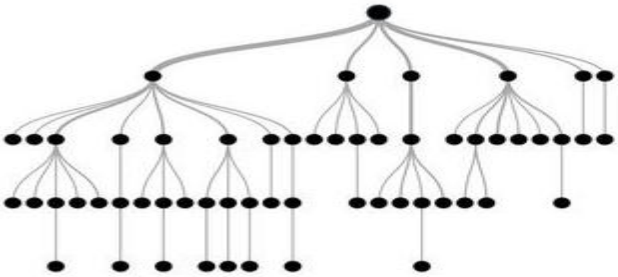

Figure.1 Decision Tree
Source: A Beginner's Guide to Decision Trees (deeplearning4j.org)

4.1.1 Regression trees

A regression tree is a type of decision-making tree. Regression tree is created using binary algorithm division, which is a collaborative unblocked technique which divides information into division or branches, and then splits each partition into several smaller groups as a result of the spreading of each branch.

First of all, all records of the training set are classified into a definite division. The rule then begins to allocate information in the first 2 branches or divisions using every available binary partition on each region. The rule selects the partition which reduces the sum of squared deviations from the interval of two different divisions at intervals. This split rule applies to each new partition. This technique continues till each node reaches the minimum node size specified by the user and becomes a terminal node [8].

Algorithm for Building a Regression Tree

1. Apply a recursive binary partition to increase an external tree on training information, preventing only each terminal node, less than the minimum number of observations.
2. Apply value complexity to big tree to pruning so that the sequence of best sub-trees can be obtained in the form of α.
3. Now apply cross-validation of K-fold to reside on α. which split the training observations into K folds. for every k = 1,2, . . .,K:
    - i. Perform Steps 1 and a pair of on virtually the kth fold of the training knowledge.
    - ii. Judge the mean square prediction error on the information within the left-out k^{th} fold, as a perform of α.

Average the results for every value of α, and choose α to reduce the average error.

4. Return the sub tree from Step 2 that corresponds to the chosen value of α.

For each value of α there corresponds a sub tree $T \subset T_0$ such

$$\sum_{m=1}^{|T|} \sum_{i:\, x_i \in R_m} (y_i - \hat{y}R_m)^2 + \alpha |T| \qquad (4.1.1)$$

is as small as potential. Here | T | the number of terminal nodes of tree T, Rm indicates that the parallelogram (i.e. the set of prediction space), and the Y Rm, similar to the mth terminal node, is that the expected response to Rm, training overview means in Rm. The calibration parameter α controls a trade-off between the sub tree's complexness and its suitable the training knowledge. When α = 0, then the sub tree T can merely equal T_0, as a result of then Eq. (4.1.1) Just measures coaching error. However, in the form of α growth, there is a value to get trees with many terminal nodes, hence the eq. (4.1.1) can be reduced to a small sub-tree. It seems that as increase α from zero in Eq. (4.1.1), branches get rich crop from tree in highly nesting and estimated fashion, so it is easy to get complete sequence of sub-trees in the form of demonstration of α. It chooses the value of α that employs the verification set or uses cross-verification. Then come to the Total Knowledge Set and produce sub-tree like α [9].

Now fit a regression tree to the chronic_kidney_disease knowledge set. First, produce a training set, and work the tree to the training knowledge. solely four of the variables (hemo = hemoprotein in gms, rc = Red blood corpuscle Count in millions/cmm, sg = relative density, bgr = blood sugar Random in mgs/dl) are utilized in constructing the tree. Within the context of a regression tree, deviance may be a measure of goodness of work of a model. Higher numbers perpetually indicates unhealthy work. The null deviance shows however well the response variable is expected by a model that features solely the intercept (grand mean) wherever as residual with inclusion of freelance variables i.e. merely the total of square errors for the tree (Residual mean deviance: 0.02417 = 4.713 / 195). Currently plot the tree that relies on prediction of ckd/notckd.

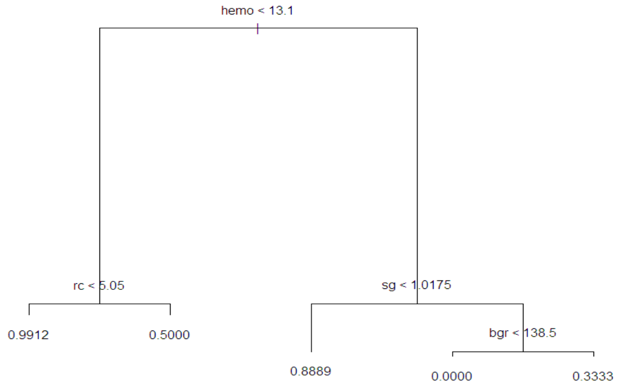

Figure.2 Regression Tree: Prediction of ckd/notckd

The variable Hb (hemo<13.1) indicate the patient is also stricken by CKD. The tree indicates that lower values of hemo correspond to a lot of likelihood of CKD. The tree predicts a particular Gravity (sg) of 0.8889 for the patient who might not be stricken by CKD (sg>1.0175 and hemo<13.1).

In keeping with the cross-validation results, it uses the unpruned tree to create predictions on the check set. The check set MSE or means that square deviation (MSD) of associate figurer measures the typical of the squares of the errors or deviations i.e., the distinction between the figurer and what's estimated related to the regression tree is 0.034. The square root of the MSE is thus around 0.184, indicating that this model leads to test predictions that are at intervals around 0.184of truth category value for a patient for not being the patient CKD.

4.1.2 Classification trees

A Classification tree records, labels, and assigns variables to discrete partitions. A Classification tree can also give evaluation of confidence that the classification is correct. A Classification tree is built from beginning to end, a process known as binary recursive partitioning. This is a step by step process of splitting the data into partitions, and then splitting it up advance on each of the branches [10]. Classification error rate is given by:

$$E = 1 - \underset{k}{max}\,(\hat{p}nk) \qquad (4.1.2)$$

Here are the $\hat{p}nk$ category which represents the proportion of training observation within the k^{th} area. However, it seems that classification is not sufficiently sensitive to the growth of the tree, and sees that 2 alternative works are given priority.

The Gini index is given as:

$$G = \sum_{k=1}^{K} \hat{p}nk(1 - \hat{p}nk) \qquad (4.1.3)$$

measure of variance thereon, the K categories. It's not onerous to check that the Gini index accept on a small value if all of the $\hat{p}nk$'s are near to zero or one. For this reason the Gini index is mentioned as a degree of node purity—a little bit value indicates that a node contains preponderantly observations from one category.

An option of the Guinee Index is cross-entropy, which is given as-

$$D = -\sum_{k=1}^{K} \hat{p}nklog\hat{p}nk \qquad (4.1.4)$$

Since $0 \leq \hat{p}nk \leq 1$, it follows that $0 \leq -\hat{p}nk\log\hat{p}nk$. One will show that the cross-entropy can closer to zero if the $\hat{p}nk$'s is close to zero or close to one. Therefore, like Gini index, the cross-entropy can take a small value if the n^{th} node is pure. In fact, it seems that the Gini index and also the cross-entropy are quite similar.

When building a classification tree, either the Gini index or the cross entropy is usually the standard accredited judge of a selected partition, because these two approaches square measure extra sensitive for node purity, this is classification error rate.

Any of those three approaches may be used once pruning the tree, however the classification error rate is preferred if prediction accuracy of the ultimate pruned tree is that the goal.

It initial use classification trees to investigate the CKD knowledge set. In this knowledge, category may be a binary variable, that takes on a value of notckd if the category variable is up to 1, and takes on a value of ckd otherwise. For classification trees it's measured that the training error rate is 1.5 attempt to the deviance reported within the output is given by-

$$\text{Training Error Rate} = -2\sum_{n}\sum_{k} x_{nk} \log \hat{p}_{nk} \qquad (4.1.5)$$

Where x_{nk} is that the variety of observations within the n^{th} terminal node that belong to the kth category. A slight deviance indicates a tree in which decent (training) knowledge is appropriate. Residual mean deviance according is just the deviance divided by n−|T0|, that during this case is 400−08 = 392.

One of the most important properties of the tree is that they will be graphically displayed in Figure 3.

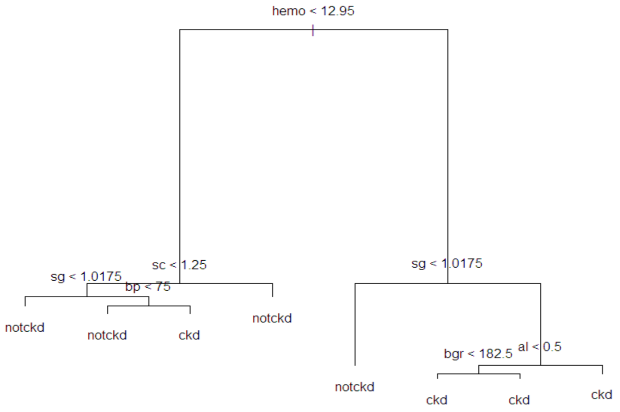

Figure.3 Classification Trees view of ckd dataset

```
node), split, n, deviance, yval, (yprob)
      * denotes terminal node

 1) root 400 529.300 ckd ( 0.625000 0.375000 )
   2) hemo < 12.95 229  48.130 ckd ( 0.978166 0.021834 )
     4) sc < 1.25 42  30.660 ckd ( 0.880952 0.119048 )
       8) sg < 1.0175 28   0.000 ckd ( 1.000000 0.000000 ) *
       9) sg > 1.0175 14  18.250 ckd ( 0.642857 0.357143 )
        18) bp < 75 7   0.000 ckd ( 1.000000 0.000000 ) *
        19) bp > 75 7   8.376 notckd ( 0.285714 0.714286 ) *
     5) sc > 1.25 187   0.000 ckd ( 1.000000 0.000000 ) *
   3) hemo > 12.95 171 145.800 notckd ( 0.152047 0.847953 )
     6) sg < 1.0175 24   8.314 ckd ( 0.958333 0.041667 ) *
     7) sg > 1.0175 147  29.290 notckd ( 0.020408 0.979592 )
      14) al < 0.5 142  11.900 notckd ( 0.007042 0.992958 )
        28) bgr < 182.5 137   0.000 notckd ( 0.000000 1.000000 ) *
        29) bgr > 182.5 5   5.004 notckd ( 0.200000 0.800000 ) *
      15) al > 0.5 5   6.730 notckd ( 0.400000 0.600000 ) *
```

The most important sign of result appears to be hemo location, because the first branch seperates ckd locations from notckd locations. R shows output for each branch of the tree. R shows the split criterion (e.g. sc<1.25, sg<1.0175, bp<75, al<0.5, bgr<182.5), the number of observations in the branch, deviation, the overall prediction for the branch (ckd or notckd), and the portion of observations in that branch which takes the values of the ckd and the notckd. Branches going towards the terminal nodes are indicated using astricks.

In order to properly evaluate the performance of the classification tree on these data, it should be guessed by the trial error instead of computing the training error. When splitting into a training set and a test set, build a tree using a training set, and evaluate its performance on test data. This approach leads to accurate predictions for approximately 93% of the test data set (119 + 67/200 = 0.93) of the locations.

Which can be find as-

```
          Result.test
Tree.pred ckd notckd
   ckd    119     4
  notckd  10     67
```

Now, it absolutely was think about whether or not pruning the tree may result in improved results. To perform this cross-validation is employed to work out the optimum level of tree quality; cost complexity pruning is employed so as to pick a sequence of trees for thought. so as to point that the classification error rate to guide the cross-validation and pruning method, instead of the default that show deviance solely, reports the amount of terminal nodes of every tree thought-about (size) also because the corresponding error rate and also the value of the cost-complexity parameter.

```
$size
[1] 5 3 2 1

$dev
[1]  7  4 12 79

$k
[1] -Inf  0.5  7.0 68.0


$method
[1] "misclass"

attr(,"class")
[1] "prune"         "tree.sequence"
```

dev matches the cross-verification error rate in this example. The tree with 4 terminal nodes returns results in the lowest cross-verification error rate with 4 cross-verification errors. As shown in the form of a function of both the size and the error rate (figure 4).

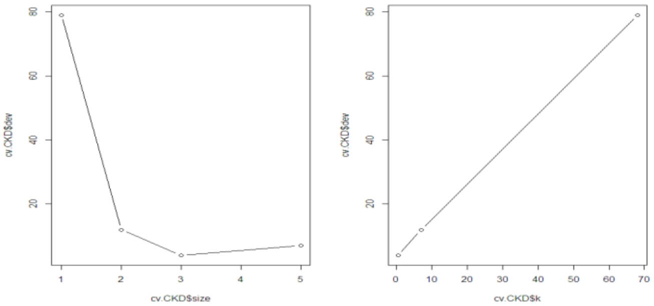

Figure 4. Error rate function of size and k

Now prune the tree to the four-node tree (figure 5).

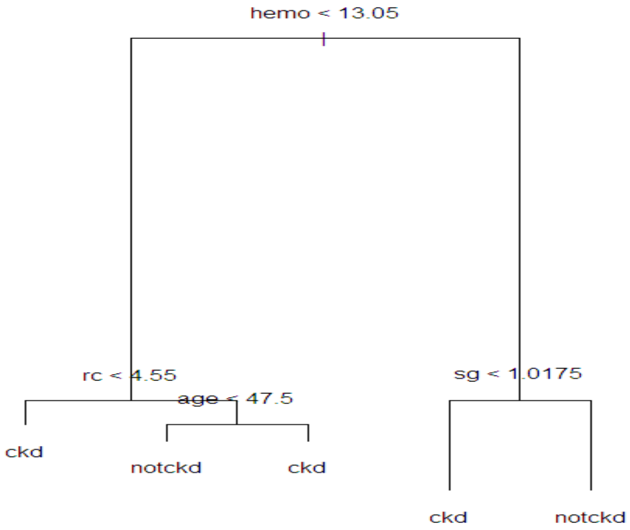

Figure 5. Prune tree view of four-node tree

5. MEDICAL SIGNIFICANCE

After analyzing the decision tree generated by regression and classification algorithms, important parameters were known supported the high strength rules. the kind of initial diagnosing compete the foremost vital role within the rules generated by each data-mining algorithms. the most causes of CKD area unit diabetes and high pressure. Studies show that patients (with the age > 47.5) with red blood cell count (rc < 4.55) have additional likelihood of CKD. On the opposite positive facet, the specific gravity (sg>1.0175) for a patient having not ckd,. under traditional hemoglobin levels indicate anemia. Anemia normally happens in individuals with chronic kidney disease (CKD)—the permanent, partial loss of kidney functions. Anemia would possibly begin to develop within the early stages of CKD, when someone has 20 to 50 % of traditional kidney perform. Anemia tends to worsen as CKD progresses. Someone has kidney failure once he or she wants a kidney transplant or dialysis so as to measure [11].

6. CONCLUSION

Although many research papers already published on the role of data mining in healthcare, but here for the instant and accurate result, a data mining approach is used to predict the disease CKD and using a popular data mining algorithm the regression and decision tree. The CKD dataset all the attributes have their usual meaning but in data mining it observed which attribute is a key to decide the particular patient suffering or not from CKD. By applying Decision tree method in regression tree model hemoglobin (hemo) is a key element and if it is less than standard result it should be subject for further investigation i.e. sg, bgr etc.

Again in classification method also the important result appears from attribute hemo. The corresponding model display the split criteria ($sc<1.25$, $sg<1.0175$, $bp<75$, $al<0.5$, $bgr<182.5$) which have their usual meanings. For the properly evaluation of the performance of a classification tree on these information, it ought to be estimate the check error instead of merely computing the training error. the complete information split observations into a training set and a check set to boost the proper prediction, build up the tree using the training set, and evaluated its performance on the check information. On the test dataset it corrects prediction for around 93% which is up to the mark than training dataset.

7. FUTURE RESEARCH

Although there are immense possibilities are present in prediction of diseases through data mining and its applications. In this sequence further work can be done in future by moving forward. Although the dataset that has been used here has a very small size, but if a large dataset is used, it can mean meaningfulness. One problem is also that the data is incomplete. In further experiments, we can access the key element of a large database that can prove the meaning of this experiment. For example, in the above experiment, the key element is hemoglobin; similarly the other attributes can be determined.

REFERENCES

[1] World Kidney Day: Chronic KidneyDisease. http://www.worldkidneyday.org/faqs/chronic-kidney-disease/, 2015.

[2] Jha V, Garcia-Garcia G, Iseki K, et al. *Chronic kidney disease: global dimension and perspectives. Lancet.* Jul 20 2013; 382(9888):260-272.

[3] Couser WG, Remuzzi G, Mendis S, Tonelli M. The contribution of chronic kidney disease to the global burden of major non communicable diseases. Kidney Int.; 80(12):1258-1270; Dec 2011.

[4] World Health Organization: Preventing Chronic Disease: A Vital Investment. Geneva, WHO, 2005.

[5] PAKDD Workshop, Toward the Foundation of Data Mining, Taipei, Taiwan,Website:www.mathcs.sjsu.edu/faculty/tylin/pakdd; workshop.html; Accessed on 2002, April 30.

[6] Fayyad U., Piatetsky-Shapiro G., Smyth P., Uthurusamy (Eds.) R., Advances in Knowledge Discovery and Data Mining, MIT Press, Cambridge, MA, 1997.

[7] James G, Witten D., Hastie T., Tibshirani R., An Introduction to Statistical Learning with Applications in R, Springer Science+Business Media, New York , 2013, www.springer.com, ISBN 978-1-4614-7138-7 (eBook).

[8] Rokach, Lior; Maimon, O. (2008). Datamining with decision trees: theory and applications. World Scientific Pub Co Inc. ISBN 978-9812771711.

[9] Quinlan J.R., (1992),. Learning with continuous classes. In Proceedings AI'92(Adams & Sterling, Eds), 343-34, Singapore: world scientific, 1992.

[10] Alsabti K., Ranka S., and Singh V. CLOUDS: A Decision Tree Classifier for Large Datasets. In Proc. of the 4th Intl. Conf. on Knowledge Discovery and Data Mining pages 2–8, New York, NY, August 1998.

[11] Brugnara C, Eckardt KU. Hematologic aspects of kidney disease. In: Taal MW, ed. Brenner and Rector's The Kidney. 9th ed. Philadelphia: Saunders; 2011: 2081–2120.

[12] Chaurasia, V., Pal, S., & Tiwari, B. B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, 12(2), 119-126.

[13] Chaurasia, V. and Pal S. Performance analysis of data mining algorithms for diagnosis and prediction of heart and breast cancer disease Rev Res 2014; 3: 1–13