International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Assessment of Chronic Kidney Disease using Classification Algorithms

Feiroz Khan T.H., Pavan Sundar Reddy G, Sai Surya Harsha D, Gopala Krishna M,

.

***Abstract:** Kidney Disease (CKD) implies the condition of kidney risk which may even get worse by time and by referring the factors. If it continues to get worse Dialysis is done and worst-case scenario it may lead to kidney failure (End-Stage Renal Disease). Detection of CKD in an early stage could help in sorting out the complications and damage. In the previous work classification used are SVM and Naïve Bayes, it resulted that the execution time took by Naïve Bayes is minimal compared to SVM, incorrect instances are less for SVM that results in less classification performance of Naïve Bayes, because of slight accuracy difference. It can be rectified by taking a smaller number of attributes. Naïve Bayes is a probabilistic classifier a simple computation by applying Bayes Theorem with a conditional independence assumption. The work mainly results in increasing diagnostic accuracy and decrease diagnosis time, that is the main aim. An attempt is made to develop a model evaluating CKD data gathered from a particular set of people. From the model data, identification can be done. This work has engrossed on developing a system based on classification methods: SVM, Naïve Bayes, KNN.*

***Keywords:** Chronic Kidney Disease (CKD), SVM, KNN, Naïve Bayes.*

I. INTRODUCTION

Chronic Kidney Disease or (CKD) refers to a condition where kidneys lose their functionality over time. If the damage is severe, it will lead to ENRD. This is called kidney failure. If kidneys fail, the alternative like Kidney replacement or Dialysis can be opted. The glomerular filtration rate (GFR) is the best pointer of how well the kidneys are functioning .

CKD has got no cure but it can be treated based on symptoms to reduce complications and decrease the rate of progression Sickness, Vomiting, Loss of hunger, Fatigue and shortcoming, Sleep issues

Manuscript published on November 30, 2019.
* Correspondence Author

Feiroz Khan*, computer science department, SRM UNIVERSITY, Chennai, INDIA. Email: feirozkh@srmist.edu.in

Pavan Sundar Reddy, computer science department, SRM UNIVERSITY, Chennai, INDIA. Email: pavansundar1999@gmail.com

Sai Surya Harsha D, computer science department, SRM UNIVERSITY, Chennai, INDIA. Email: dssharsha99@gmail.com

Gopala Krishna M, computer science department, SRM UNIVERSITY, Chennai, INDIA. Email: gk1729@outlook.com

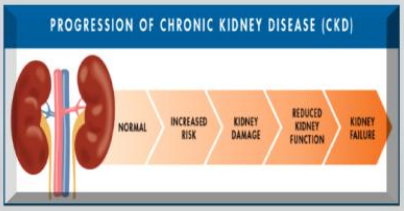


Fig-1 : Progression of CKD

Changes in the amount you urinate. These are basic symptoms of CKD. The term GFR (Glomerular Filtration Rate) is the important and best test to measure how well kidney function. The GFR rate decrease as with the age, so if GFR is less than 60 then its is considered the patient is suffering from CKD.

S.No	Stages of CKD	Glomerular Filtration Rate (GFR)	Action Plans
1	Kidney damage with normal GFR	90 or above	Diagnosis and treatment of comorbid conditions, disease progression, reduction of risk factors for cardiovascular disease
2	Kidney damage with mild decrease	60 to 89	Estimation of disease progression
3	Moderate decrease	30 to 59	Evaluation and treatment of disease complications
4	Severe reduction	15 to 29	Preparation of kidney replacement treatment (dialysis, transplantation)
5	Kidney failure	Less than 15	Kidney replacement therapy

Fig-2 : GFR range

II. LITERATURE SURVEY

Abhay Bansal , A.Sai Sabitha , Kushboo , Chandel and Veenitha Kunwar et.al 1 ,(2016) 'Chronic Kidney Disease Analysis Using Data Mining Classification Techniques',IEEE,used Data Mining classification techniques such as Naïve Bayes and Artificial Neural Network(ANN) for predicting CKD.

The experimental results implemented in RapidMiner showed that Naïve Bayes produce accurate results than Artificial Neural Network.

Abhinandan Dubey et.al 2, (2015) 'A Classification of CKD Cases Using Multivariate KMeans', International Journal of Scientific and Research Publications, Volume 5, Issue 8,adopted the K-means clustering algorithm with a single mean vector of centroids, to classify and make clusters of varying probability of likeliness of suspect being prone to Chronic Kidney Disease.The results were obtained from a real case dataset from UCI machine learning repository.

Anusorn Charleonann, Nitat nichawee, Sathit Suwannawach, Thipwan Fufaung, Tippawan Niyomwong and Wandee Chokchueypattanakit et.al 3, (2016) 'Predictive analytics for Chronic Kidney Disease using Machine Learning Techniques', The 2016 Management and innovation Technology International Conference. used machine learning techniques such as KNN, SVM , Logistic regression and Decision Tree. These predictive models were constructed from chronic kidney disease dataset and the performance of these models were compared together in order to select the best classifier for predicting the CKD.

Azamat Serek,Sahhriar Shamiluulu and Yedhilkhan Amirgaliyev et.al 4, (2015) 'Analysis of Chronic Kidney Disease Data set by Applying Machine Learning Methods', JSCE International journal of Computer science Volume - 3,issue-8,used Support Vector Machine Algorithm for finding the effects of using clinical features to classify patients with chronic kidney disease. The CKD dataset was based on clinical history, physical examinations and laboratory tests.

Che-Lun Hung,Chuan Yi Tang,Hanyu Zhang and William Cheng-Chung et.al 5,(2018) 'Chronic Kidney Disease survival prediction with Artificial Neural Networks',IEEE, used data preprocessing, data transformations and Artificial Neural Networks(ANN) to establish the mapping from many clinical factors to the patients survival.

Derya Avci,Engin Avci,Songul Karakus and Ozlem Ozmen et.al 6,(2018) 'Performance comparison of some classifiers on Chronic Kidney Disease data.',IEEE., using WEKA software,the dataset was classified according to whether it is chronic kidney disease using Naïve Bayes,K-star,SVM and J48 classifiers in data mining.Accuracy,precision,sensitivity and F-measure values are used for performance comparisons of the performed classifications.

DiniWidianti,MubarikAhmad,Peny Amalia,VitriTunjungsari and Ummi Azizah Rachmawati et.al 7 ,(2016) 'Diagnostic Decision Support System of Chronic Kidney Disease using Support Vector Machine',IEEE ,developed a decision support system for a doctor in diagnosing of the kidney disease patients. The system displays the results of predicting whether the patients with renal disease have entered a phase of chronic kidney disease or not.

D.M. Perera K.A.D.C.P Kahandawaarachchi and M.P.M.N Wickramasinghe et.al 8,(2017) 'Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms. 'IEEE,performed different algorithms like Multiclass Decision Jungle,Multiclass Decision forest, Multiclass Neural Network and Multiclass Logistic regression. The experimental results showed that Multiclass Decision Forest gives a better result than the other algorithms.

Dr.R.Thirumalaiselvi and S.Dilli Arasu et.al 9 ,(2018) 'Review of Chronic Kidney Disease based on Data Mining Techniques', International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 ,major data mining techniques such as clustering,classification,association analysis,regression,time series and sequence analyses were used to predict CKD.The techniques that were introduced had minor drawbacks in the quality of preprocessing or at any other stages.

Dr.Uma N Dulhare and Mohammad Ayesha et.al 10 ,(2016) 'Extraction of action rules for chronic Kidney Disease using Naïve Bayes classifier',IEEE ,predicted CKD using Naïve Bayes with one R attribute selector which helps to prevent the advancing of chronic renal disease for stages and extracted action rules based on stages.

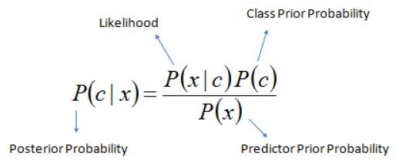
N.Radha and S.Ramya et.al 11 ,(2018) 'Performance Analysis of Machine Learning Algorithms for Predicting Chronic Kidney Disease', International Journal Of Management, Technology and Engineering ,used machine learning algorithms for classifying different stages of CKD,experimental results were performed on different algorithms like Naïve Bayes, Decision Tree,KNN and SVM.

III. METHODOLOGY

A. NAÏVE BAYES:

Given the data b plus any initial knowledge about the prior probabilities of the various hypotheses in a. This Classifier comes under the group of probabilistic classifiers.

Bayes' Theorem is stated as:



$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

B. DECISION TREE:

A decision tree is a classifier as a tree structure with two sorts of hubs or nodes: Decision hub: Specifies a decision or trial of some characteristic, with one branch for every result. Leaf hub: Indicates order of a model

C. LOGISTIC REGRESSION

Logistic Regression is a system in Machine Learning, most likely it's a measurable investigation technique used to anticipate an information worth dependent on earlier perceptions of an informational index. The logisitic capacity is additionally called sigmoid capacity, it brings about a S-formed bend, it can take any genuine esteemed numbers or something like that and is mapped into values 0 or 1.

1 / (1 + e^-value)

D.SVM:

Support vector machine comes under supervised learning and generally, it is defined by a separating hyperplane. SVM is a very popular and useful classification for good computations and large data. SVM has got a clever way to prevent overfitting because they use many features without requiring too much computation. In SVM, support vectors are defined and from this, we find a margin with minimum length and also keeping the constraint so that features have good confidence from the plane. The margins cane be 1) functional margin 2) geometrical margin.

E.K NEAREST NEIGHBOUR

Occurrence based learning is One method for fathoming assignments of approximating discrete or genuine esteemed objective capacities and KNN is one this sort of grouping. In KNN not the same different characterizations, we store preparing Examples, when a test model is given we discover the closest neighbor.

At prediction time:

$$Dist(c_1, c_2) = \sqrt{\sum_{i=1}^{N} \left(attr_i(c_1) - attr_i(c_2)\right)^2}$$

$$k - NearestNeighbors = \left\{k - MIN(Dist(c_i, c_{test}))\right\}$$

$$prediction_{test} = \frac{1}{k}\sum_{i=1}^{k} class_i \ (or \ \frac{1}{k}\sum_{i=1}^{k} value_i)$$

IV. WORK DONE

By utilizing this model, can capable foresee whether the individual is having CKD or not. This depends on the highlights of the patient like RC, WC, BGR and so on. By anticipating this can say that the individual is affected or not. In the event that he/she is affected we can offer some guidance for further procedures. We present the prescient models by utilizing AI strategies including K-closest neighbors (KNN), support vector machine (SVM), calculated relapse (LR), and choice tree classifiers to foresee incessant kidney illness. From the test results, it tends to be seen that SVM classifier gives the most elevated exactness. Moreover, SVM has most elevated affectability subsequent to preparing and testing by the proposed technique. Subsequently, it tends to be presumed that SVM classifier is appropriated for foreseeing the interminable kidney ailment. Comparion study of clasiification algorithm.

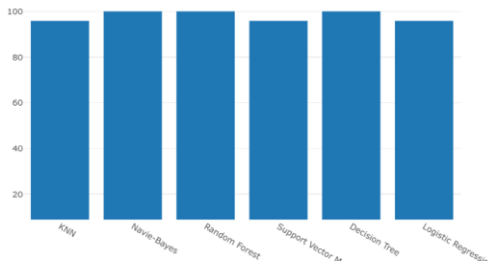


Fig-3 : Comparison of Classification Algorithm.

V. FUTURE SCOPE

There is a need of an indicative choice emotionally supportive network to improve specialist's judgment in deciding an interminable state of kidney sickness' patients. We proposed a framework that will help specialists treat their kidney illness' patients. This framework was created dependent on AI methods. The strategy of this investigation is partitioned into two fundamental stages: grouping displaying and framework advancement. From arrangement demonstrating, we discovered principles and model in the grouping of kidney sickness. Moreover, we executed that guidelines into a framework utilizing programming language python. This analysis choice emotionally supportive network has some fundamental highlights, for example, Read File, Classification, Error Rate, and Diagnosis. This framework is relied upon to help the specialist in choosing a constant state of kidney malady patients with desired accuracy.

VI. RESULTS

By using this model, can able predict whether the person is having CKD or not. This is based on the features of the patient like RC, WC, BGR etc. By predicting this can say that the person is effected or not. If he/she is effected we can give some advice for further proceedings.

RANDOM FOREST: Random forests is a notion of the general technique of random decision forests that are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees, the accuracy of the algorithm is 100%.

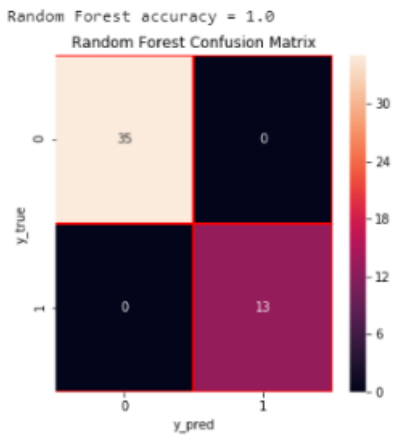


Fig-4 : Random Forest Accuracy Matrix

K NEAREST NEIGHBOUR In pattern recognition, the k-Nearest Neighbours algorithm (or k-NN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space, the accuracy of the algorithm is 95.8%.

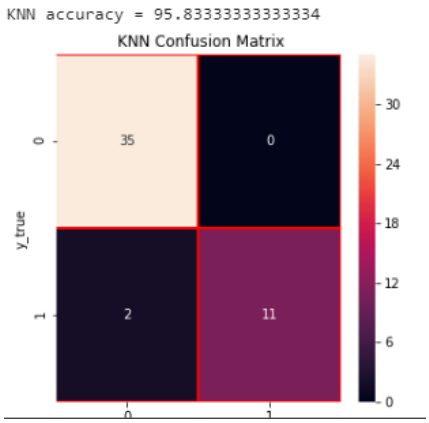


Fig-5 : KNN Accuracy Matrix

SVM: In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis, the accuracy of the algorithm is 95.8%.

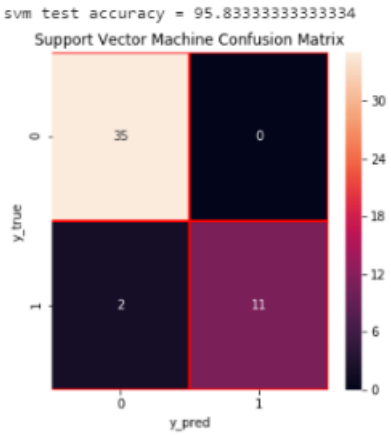


Fig-6 : SVM Accuracy Matrix

NAÏVE BAYES: In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features, the accuracy of the algorithm is 100%.

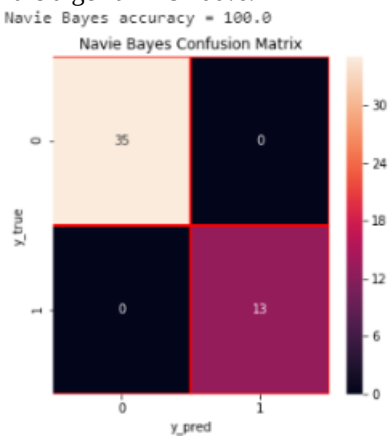


Fig-7 : Naïve Bayes Accuracy Matrix

DECISION TREE: A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm, the accuracy of the algorithm is 100%.

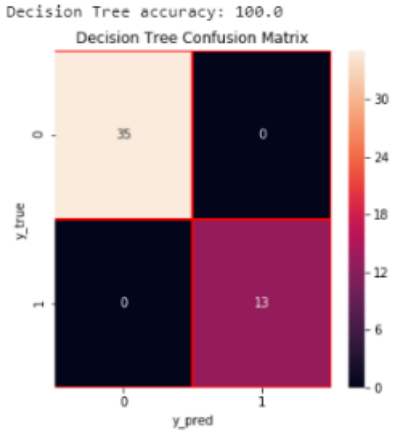


Fig-8 : Decision Tree Accuracy Matrix

LOGISTIC REGRESSION: In statistics, logistic regression, or logit regression, or logit model is a Regression model where the dependent variable (DV) is categorical, the accuracy of the algorithm is 95.8%.

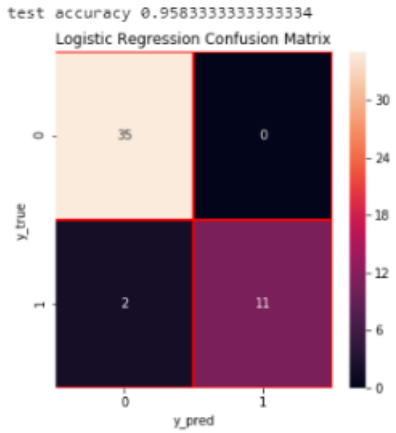


Fig-9 : Logistic Regression Accuracy Matrix

VII. CONCLUSION

In this work, we have done prediction models by using some of the machine learning algorithms such as Support Vector Machine, K Nearest Neighbor, Logistic Regression, Naïve Bayes, Decision Tree and through the trained model, we predict whether the person has CKD or no. From the results, it can be seen that Naïve Bayes Classification algorithm is having more accuracy than SVM, Regression and Clustering algorithms. NAÏVE BAYES results as a most accurate classifier because it has least compile time and works efficiently with less number of test attributes. And its also a important factor for this model to show fast results .

REFERENCES

1. Abhay Bansal, A.Sai Sabitha,Kushboo and Chandel, Veenitha Kunwar,(2016) 'Chronic Kidney Disease Analysis Using Data Mining Classification Techniques',IEEE.
2. Abhinandan Dubey,(2015) 'A Classification of CKD Cases Using MultiVariate KMeans', International Journal of Scientific and Research Publications, Volume 5, Issue 8.

3. Anusorn Charleonann,Nitat nichawee,Sathit Suwannawach,Thipwan Fufaung,Tippawan Niyomwong and Wandee Chokchueypattanakit,(2016) 'Predictive analytics for Chronic Kidney Disease using Machine Learning Techniques', The 2016 Management and innovation Technology International Coference.
4. Azamat Serek,Sahhriar Shamiluulu and Yedhilkhan Amirgaliyev,(2015) 'Analysis of Chronic Kidney Disease Data set by Applying Machine Learning Methods', JSCE International journal of Computer science Volume -3,issue-8.
5. Che-Lun Hung,Chuan Yi Tang,Hanyu Zhang and William Cheng-Chung,(2018) 'Chronic Kidney Disease survival prediction with Artificial Neural Networks',IEEE.
6. Derya Avci,Engin Avci,Songul Karakus and Ozlem Ozmen,(2018) 'Performance comparison of some classifiers on Chronic Kidney Disease data.',IEEE.
7. Dini Widianti,Mubarik Ahmad,Peny Amalia,VitriTunjungsari and Ummi Azizah Rachmawati,(2016) 'Diagnostic Decision Support System of Chronic Kidney Disease using Support Vector Machine',IEEE.
8. D.M. Perera K.A.D.C.P Kahandawaarachchi and M.P.M.N Wickramasinghe,(2017) 'Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms.',IEEE.
9. Dr.R.Thirumalaiselvi and S.Dilli Arasu ,(2018) 'Review of Chronic Kidney Disease based on Data Mining Techniques', International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23.
10. Dr.Uma N Dulhare and Mohammad Ayesha,(2016) 'Extraction of action rules for chronic Kidney Disease using Naïve Bayes classifier',IEEE.
11. N.Radha and S.Ramya,(2018) 'Performance Analysis of Machine Learning Algorithms for Predicting Chronic Kidney Disease', International Journal Of Management, Technology and Engineering.

AUTHORS PROFILE

T.H.Feiroz Khan, received his M.E. degree in Computer science Engineering from Annamalai University, Tamilnadu in 2004. Currently pursuing PhD in school of computing in the Domain of Wireless Networking. He is a currently Assistant Professor in SRM Institute of Science and Technology, Chennai. He has teaching work experience of 15 years in the Engineering and Technology. He is member of International Association of Engineers (IAENG).His research interest includes Wireless System and Machine Learning.

G. Pavan Sundar, Reddy has received his Board of Intermediate Education certification from Sri Chaitanya College, Vijayawada in 2017. Currently pursuing UG in Computer science and engineering from SRM institute of Science and technology, Chennai. . His research interest includes Machine learning

Darbhamalla Sai Surya, Harsha has received his Senior school Certificate Examination of the central Board of secondary education (CBSE) held in March,2017 from Madhava Kripa School, Manipal, Karnataka and is currently pursuing BTech in SRM Institute of Science and Technology Ramapuram, Chennai. His research interest includes Machine learning

Gopala Krishna, .M received has received his Board of Intermediate education certificate held in March,2017 from Narayana IIT Academy, Hyderabad, Telangana. He is Currently pursuing UG in Computer science and engineering from SRM institute of Science and technology, Chennai. His research interest includes Machine learning