International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Prediction of Diagnosing Chronic Kidney Disease using Machine Learning: Classification Algorithms

M. Prashanthi Reddy, T. Uma Devi

Abstract: ***Chronic Kidney Disease is a very dangerous health problem that has been spreading as well as growing due to diversification in life style such as food habits, changes in the atmosphere, etc. The branch of biosciences has progressive to a bigger extent and has bring out huge amounts of data from Electronic Health Records. The primary aim of this paper is to classify using various Classification techniques like Logistic Regression (LR), K-Nearest Neighbor (KNN) Classifier, Decision Tree Classifier Tree, Random Forest Classifier, Support Vector Machine (SVM), and SGD Classifier. According to the health statistics of India 63538 cases has been registered on chronic renal disorder. Average age of men and women susceptible to renal disorders occurs within the range of 48 to 70 years.***

Keywords: Chronic kidney Disease, logistic Regression, K-Nearest Neighbor, Support vector machine (SVM).

I. INTRODUCTION

Machine learning and data processing plays a vital role in getting more flexible and understandable reports on the idea of varied techniques. Kidneys role act as blood purifiers that remove waste contents while preserving new valuable blood contents like proteins. If the purifiers were damaged, the protein content will be initially leaked and the substances may seep into urine from blood. Sometimes chronic renal disorder is amid high vital sign, which not only are often caused by kidney damage but also further accelerates kidney injury and may be a major reason for the negative effects of chronic renal disorder on other body parts, automatically increases risk of heart condition and heart-strokes, collection of excess body fluids, anemia, weakening of bones and deterioration mainly the body will not support for medications. Chronic kidney disease develops slowly, with few symptoms. It can't be detected until the seriousness of the disease is advanced. If it's detected early, treatment can hamper or refrain kidney function deny and reduces the contradictory effects on new body parts. A biopsy measuring tool called glomerular filtration rate works on the kidneys for removing waste blood contents called creatinine. If the value lies within the range of 60 to 90 it be an early sign of occurring kidney disease; a worth below 60 is typically considered as abnormal phase.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
M. Prashanthi Reddy*, PG Student, Department of CS, GIS, GITAM (Deemed to be University), Visakhapatnam, India, mallepalliprashanthi@gmail.com
Dr. T.UmaDevi, Associate Professor**,** Department of CS, GIS, GITAM (Deemed to be University), Visakhapatnam, India, umadevi.tatavarthi@gitam.edu

By testing of urine samples it gives the results of protein contents (albumin) within the urine; repeated results of 30 mg or more can signify a drug. Huge vital signs can also point to underlying chronic renal disorder. Distinct machine learning procedures are appropriate for analyzing the data from distinct prospects and reviewing it into beneficial data.

II. MACHINE LEARNING ALGORITHMS– CLASSIFICATION

A. K-Nearest Neighbors Classifier Algorithm

The K-nearest neighbors is a Classification technique and it is one of the most important classification techniques in machine learning. KNN belongs to the supervised learning domain and have various operations in pattern recognition, Processing and intrusion detection. Making of prognosis for a replacement datum, the data discovers the nearest neighbors within the training data set. By giving the previous data, the KNN segregates the coordinates into groups that are classified by a selected aspect.

B. Logistic Regression Algorithm

Logistic Regression is a Classification technique used for assigning observations into the discrete arrangement of classes. In general, Rectilinear Regression and Logistic Regression are very much alike. Linear Regression techniques are utilized for forecast the values where as Logistic Regression is employed for Classification tasks. In Logistic Regression, rather for fitting a regression curve, fitting of "S" shaped logistic function predicts two utmost values (0 or 1). Logistic Regression is often utilized for segregating the observations using various sorts of data and may easily conclude the foremost competent variables utilized for the Classification.

C. Decision Tree Classifier Algorithm

Decision Tree techniques are utilized for both grouping and forecasts in AI. Utilizing the decision tree with a given arrangement of values, one can follow the different results that are an after-effect of the outcomes or choices. The decision tree is an after-effect of different various steps that will assist in reaching certain choices. To assemble decision tree, there are two stages: Induction and Pruning**.**

D. Random Forest Algorithm

Random forest algorithm is a Classification technique and it erects a various decision trees to go about as a group of arrangement and relapse process.Similarly, the random forest classifier generates huge number of trees in the forest results in high enumerate outcomes.

The Main advantage of this algorithm is reduction in over-fitting and also, in most of the cases it gives more accurate results than decision tree and it is slow in predicting real time data and difficult to implement.

E. Support Vector Machine Algorithm

Support Vector Machine (SVM) procedure is a linear model for both the classification and regression. SVM can be utilized to settle both linear and non linear issues. The fundamental thought of SVM is to locate the ideal hyper-plane between information of two classes in the preparation information.

F. Stochastic Gradient Descent (SGD) Classifier

Stochastic Gradient Descent could even be a Classification machine learning algorithms that are adept for enormous large-scale learning. SGD is a productive methodology towards discriminative learning of linear classifiers under the curved misfortune work which is linear (SVM) and logistic regression. We apply SGD to the gigantic scope AI issues that are available in content arrangement and different territories of tongue processing. It can productively scale the issue that have more than 10^5 preparing models furnished with more than 10^5 highlights. The main advantage of SGD algorithm is very efficient and will implement these algorithms quite easily. The disadvantage is that SGD calculation requires various hyper parameters such as regularization and various cycles. SGD is additionally very sensitive to include scaling, which is one of the most significant strides under data pre-processing.

III. DATA SET AND ATTRIBUTES

Experiments are directed on Chronic Kidney Disease Dataset which has been downloaded from Kaggle. This dataset contain 16 attributes (counting objective class characteristic) and 396 instances. This dataset contains information about various patients experiencing disease. The Foremost step is data preprocessing, data transformation and different classifiers to predict CKD and also proposes best forecast framework for CKD. Hence to identify the best classifier the dataset was part into two sections-Training dataset and Test dataset. Each set contains both dependent features X and an output features Y. The dataset is splitted into 75% of training data and 25% of testing data. The after effects of the system show promising consequences of better prediction at a beginning time of CKD and it is shown in below Figures.

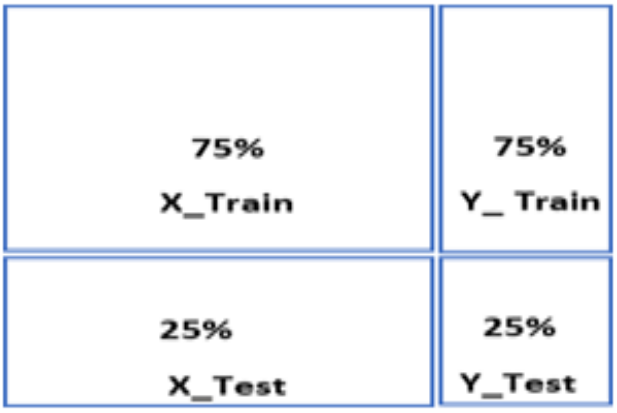

Fig.1.Representation of Training Data and Testing Data

	age	bp	sg	al	su	hemo	pcv
count	396.000000	396.000000	396.000000	396.000000	396.000000	396.000000	396.000000
mean	51.654040	75.813131	1.026904	1.255051	0.734848	12.295707	38.257576
std	16.713339	14.483353	0.066565	1.513852	1.377173	3.166023	9.321233
min	3.000000	32.000000	1.005000	0.000000	0.000000	2.600000	9.000000
25%	42.000000	70.000000	1.015000	0.000000	0.000000	10.300000	32.000000
50%	54.000000	80.000000	1.020000	1.000000	0.000000	12.500000	39.000000
75%	64.000000	80.000000	1.025000	2.000000	1.000000	14.700000	45.000000
max	90.000000	180.000000	2.025000	7.000000	8.000000	26.000000	85.000000

Fig.1. Chronic Kidney Disease Dataset

IV. RESULTS AND ANALYSIS

A comparative study on different machine learning algorithms is performed on CKD data set. The training data exactness and testing data precision of each calculation is created so as to get careful and successful outcomes concerning the informational index of CKD. The results of the algorithms are shown below:

Table- I: Comparison values

Algorithms	Training data Accuracy	Test Data Accuracy
K-Nearest Neighbors Classifier	0.74	0.69
Logistic Regression	0.75	0.75
Support Vector Machine	1.00	0.75
Random Forest	1.00	0.74
Decision Tree Classifier	0.80	0.72
SGD(Stochastic Gradient Descent) Classifier	1.00	0.75

V. CONCLUSION AND FUTURE SCOPE

It is essential to predict Chronic Kidney Disease accurately as it is stated as a deadly disease. CKD is predicted using six classifiers as of now. The Logistic Regression algorithm helps in programmed location of CKD with high exactness of 0.75. The exhibition of the models is assessed dependent on the precision of expectations. As per the outcomes shown by all the six algorithms, the accuracy of both the training dataset and testing dataset, Logistic algorithm is giving an accurate value of 0.75. As a future extension, there is a chance of applying other algorithms which are present in the classification model of the machine learning model.

REFERENCES

1. J. RadhaKrishnan, M.Sumit, "KI Reports and World Kidney Day", Kidney international reports, ISSN: 2468-0249, vol.2 (2), pp. 125-126, Mar. 2017.
2. Sossi Alaoui S., Aksasse B., Farhaoui Y. (2019) Statistical and Predictive Analytics of Chronic Kidney Disease. In: Ezziyyani M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2018). AI2SD 2018. Advances in Intelligent Systems and Computing, vol 914. Springer, Cham.
3. Sandeep Reddy Mula, Jaya. (2018). CKD Analysis Using Machine Learning Algorithms. International Journal for Research in Engineering Technology .6.3367-3379. 10.22214/ijraset.2018.1469.
4. Aljaaf, Ahmed J. et al. "Early Prediction of Chronic Renal Disorder Using Machine Learning Supported by Predictive Analytics." 2018 IEEE Congress on Evolutionary Computation (CEC) (2018): 1-9.
5. Er. Ajay Sharma , Milandeep Arora , CKD Detection by Analyzing Medical Datasets in Weka , International Journal of Computer Application (2250-1797) Volume 6– No.4, July- August 2016.
6. N.Radha , S.Ramya , Performance Analysis of Machine Learning Algorithms for Predicting CKD ,International Journal of Computer Sciences &Engineering(ICSE), Volume-3, Issue-8 E-ISSN: 2347- 2693[2015].
7. Classification Techniques:
   Hint: https://www.edureka.co/blog/classification-algorithms/

AUTHORS PROFILE

M. Prashanthi Reddy pursuing Master of Computer Applications, Department of CS, GIS, GITAM (Deemed to be University), Visakhapatnam. Her area of interest is Machine Learning and Data Mining.

Dr. T. Uma Devi is working as Associate Professor in the Department of Computer Science, GITAM (Deemed to be University), Visakhapatnam, India. She did her PhD from Acharya Nagarjuna University in 2011. Her research interests include Neural Networks, Data Mining, Bio Informatics and Data Analytics. She has been published good number of research publications in reputed international journals and guiding good number of research scholars.