

PERFORMANCE EVALUATION OF DIFFERENT MACHINE LEARNING ALGORITHMS FOR PREDICTION OF CHRONIC KIDNEY DISEASE

Muktevi Srivenkatesh

*Associate Professor, Department of Computer Science, GIS, GITAM Deemed-to-be-University, Visakhapatnam,
Andhra Pradesh, India*

ABSTRACT

Chronic kidney disease is rising health hassles and includes stipulations that minimize the efficiency of renal features and that damage kidneys. Chronic kidney sickness may be detected with countless machine learning techniques, and these have been classier. The use of a number features and classier combinations.

Methods: *In this study, we applied 12 one of a kind of machine learning classifiers (Naïve Bayes, RandomTree, REPTree, etc.) for the analysis of Chronic kidney disease. The classification performances are evaluated with five different overall performance metrics, i.e., accuracy, kappa, Mean absolute error (MAE), Root Mean square error (RMSE) and F-measures. The goal of this lookup work is to predict kidney disease with the aid of using more than one computing machine learning algorithms that are J48 Graft Decision tree (C4.5) and Bayesian Network (BN) and LMT, LAD Tree, Random Tree and Random Forest, etc.*

Results: *The machine learning algorithms under study were able to predict liver disease in patients with accuracy between 76.13% and 83.41%.*

Conclusions: *It was shown that Random forest has better Accuracy (83.41%) when compared to different machine-learning algorithms.*

KEYWORDS: *Chronic Kidney Disease, Classification, Machine Learning Algorithms, Bayesian Network, Random Fores*

Article History

Received: 13 Nov 2019 | Revised: 18 Nov 2019 | Accepted: 11 Dec 2019

INTRODUCTION

Classification is a state of information assessment that concentrates on models portraying essential insight classes. Such models, known as classifiers predict all out (discrete, unordered) class marks.

This record examination venture is an occurrence of numeric forecast, where the mannequin built predicts a consistent esteemed capacity, or requested worth, as unfriendly to an order label. This model is an indicator. Relapse examination is a factual procedure that is most normally utilized for numeric forecast.

The timespan "incessant kidney ailment (CKD)" means enduring mischief to the kidneys that can be more awful after some time. If the damage is dreadful, your kidneys may likewise quit any pretense of working. This is called kidney disappointment or quit arrange after some time condo disease (ESRD). On the off chance that your kidneys come up short,

you will require dialysis or a kidney transplant so as to live.

Anyone can get CKD. Some individuals are at extra risk than others. A few issues that broaden the chance for CKD comprises of diabetes, high-pulse, high-circulatory strain, heart malady, being more than 60 years of age. Incessant kidney infection (CKD) alludes to everyone of the five levels of kidney harm, from slight damage in stage I to entire kidney disappointment in stage V. The degrees of kidney disorder depend on how appropriately the kidneys can carry out their responsibility – to sift through waste and more noteworthy liquid out of the blood. In the early degrees of kidney illness, your kidneys are regardless ready to sift through waste from the blood. In the later degrees, the kidneys need to work more enthusiastically to dispose of waste and may furthermore quit any pretense of working.

The estimated glomerular filtration rate (eGFR) estimates how appropriately the kidneys channel waste from the blood. The degrees of waste are principally found on the eGFR assortment.

- Stage 1: Kidney illness limits kidney damage and an eGFR bigger than 90.
- Stage 2: Kidney issue limits kidney harm and an eGFR somewhere in the range of 60 and 89
- Stage 3: Kidney sickness eGFR between 30 and 59.
- Stage 4: Kidney sickness eGFR between 15 and 30.
- Stage 5: Kidney sickness in an eGFR less than 15

In this existing paper, we practice an election tree classifier (C4.5) [1], which is among the most influential information mining algorithm in research community and among the pinnacle of 10 data mining algorithms. Our goal is to predict chronic kidney sickness by gaining knowledge of algorithms.

The remaining of the research discussion is organized as follows: Section 2 briefs literature, section 3 describes brief description of selected algorithms, section 4 describes patient data set and attributes. Section 5 discusses proposed technique. Section 6 describes analysis of various algorithms. Section 7 describes performance measure of classification. Section 8 briefs discussion and evaluated results and Section-9 determines the conclusion of the research work and 10 describes References

Literature Survey

Sujata Drall, Gurudeep Singh Drall and Sugandha Singh, [2]: Chronic kidney disease (CDK) is defined by the presence of kidney damage which lasts longer than three months with decreased glomerular filtration rate (GFR). This data has been fed into Classification algorithms. The experimental results show that Naïve Bayes Algorithm gives an accuracy of 96.25%, whereas K-Nearest Neighbor came up with an accuracy of 100%.

N. Radhaand and S. Ramya [3]: Chronic kidney disease refers to the condition of kidneys caused by diabetes conditions. These problems may happen gently for a long period of time, often without any symptoms. The experimental results performed on different algorithms like Naive Bayes, Decision Tree. The experimental result shows that the K-Nearest Neighbor algorithm gives better result than preferred outcome over the other arrangement calculations and produces 98% precision.

K. R. Ananthapadmanaban and G. Parthiban [4]: On comparing the classification algorithms with respect to Naïve Bayes and Decision Tree, we came to the conclusion that the accuracy is up to 91%

N. Radha and Ramya S.,[5]: A GFR of 90 or above is considered as would be expected. Indeed, even with an ordinary GFR, it might be an expanded hazard for creating CKD if the patients have diabetes, circulatory strain in high, or a family ancestry of kidney infection.

Pavithra, N. *et al.* [6] described a symbolic fuzzy clustering algorithm with fuzzy information in the structure. The system was presented to predict and diagnose patients with renal dysfunction. The FCM clustering algorithm used to be applied to the location of the sickness in kidney disease affected person files.

Veenita Kunwar *et al.* [7] presented the prediction and diagnosis on Constant Kidney Disease utilizing information mining classifiers, example.g., ANN and Naive Bayes. The tool named as RapidMiner is used to compare the performances of both mining classifiers. The results concluded that Naive Bayes displays better accuracy (100%).

Basma Bookended *et al.* [8] discussed three learning algorithms on a set of medical data and predicted multiple machine learning algorithms that are Support Vector Machine (SVM), Decision Tree (C4.5) and Bayesian Network (BN) and chose the most efficient one.

Sharma and Rohit [9] detected and explained kidney diseases as a prelude to a suitable remedy for patients. The device was once used for identification in sufferers with kidney disease and the results of their rules expected the presence of a disease. Generally, effects primarily based on arithmetic tend to have greater accuracy.

Brief Description of Algorithms Selected for Comparison

In this section, we discuss current elements of various data mining algorithms for foregoing comparative study: Bayesian Network and Naïve Bayes Bayesian Network: A Bayesian people group is only a graphical portrayal of contingent probabilities. A implies that the possibility of B is adapted on A's worth or in math, $P(B|A)$. Guileless Bayes and Bayesian Regression can be composed as a Bayesian system.

Bayesian Inference: Bayesian Inference is the point at which we use Bayes Rule to accomplish the restrictive shot of some parameter given the information $P(Y|X)$, above. This is simply standard programming of the declaration above; however, X is taken to portray the discovered information.

Naive Bayes: Similarly to Bayesian Inference,'Naive Bayes' is just an ability we are expecting X and Y above speak the exact things in the use of Bayes Rule- - to be specific, X speaks to the element records and Y speaks to the characterization marks. Normally, we reason to discover $P(Y|X)$. The 'Naive' part originates from the presumption of autonomy between highlights.

ADTree is an altering choice tree, which is a computer studying technique for classification. It sums up choice shrubberies and has associations with boosting. An ADTree comprises of a variation of decision hubs, which determine a predicate condition, and expectation hubs, which include a solitary number. An occasion is categorized by means of an ADTree with the aid of following all paths for which all choice nodes are true and summing any prediction nodes that are traversed.

Decision Tree J48

In this practical, the general execution of choice tree J48 has been assessed and contrasted at different calculations. Arrangement calculations for the most part find a standard or set of strategies to speak to the realities and sorted into classes. The choice tree is of arrangements to speak to the realities and classified into class'. The choice tree is a well-

known basic structure that utilizes partition and vanquish method to harm down a muddled decision making process into an accumulation of straightforward choices. The choice tree component is evident and along these lines displaying an interpretable arrangement.

Given a data base $D = \{t_1, t_2, \dots, t_n\}$, where $t_i = \{t_{i1}, t_{i2}, \dots, t_{ih}\}$ and the database pattern comprises of the characteristics $\{A_1, A_2, A_3, \dots, A_h\}$. It is additionally given a lot of classes $C = \{1, \dots, m\}$.

A decision tree computational mannequin related with D that has the accompanying properties

- Each internal hub is marked with a property, A_i .
- Each arc is named with a predicate that can be applied to the property related with the parent.
- Each leaf hub is named with a class, C_j .

Given a lot of classes $C = \{1, \dots, m\}$ with equivalent probability of occurrence the entropy is $-p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots - p_m \log_2 p_m$, where p_i is the likelihood of predominance of i . Attribute with the most reduced entropy is picked as split guidelines for the tree. Tree pruning is done a base up style. It is utilized to upgrade the forecast and characterization exactness of the calculation by limiting over fitting.

J48 Graft is a calculation having purpose to grow the likelihood of grouping properly the cases. This calculation creates exclusively single tree and lessens forecast blunder. J48 join calculation is for producing united choice tree from a J48 tree calculation. The thought process of this joining calculation is to expand the probability of effectively arranging cases that fall outside the regions secured by the training information. The joining strategy is an inductive system which adds hubs to construe choice trees with the intention of lessening forecast blunders. The J48 joining calculation offers an incredible run-of-the-mill forecast precision over safeguard determination of the learning procedure.

Logistic Model

A Logistic Model Tree (LMT) fundamentally comprises of an in vogue decision tree shape with strategic relapse capacities at the leaves. The LMT comprises of a tree structure that comprises a lot of interior or non-terminal hubs and a lot of leaves or terminal hubs. The Logistic Model Tree calculation makes a tree with two-fold and multi-class target factors, numeric and missing qualities. LMT is a blend of enlistment trees and calculated relapse. LMT utilizes cost-unpredictability pruning. This calculation is obviously much slower than different calculations.

Random Tree

Random Tree (RT) is a proficient calculation for building a tree with K arbitrary angles at every hub. Random tree is a tree which is drawn indiscriminately from a lot of potential trees. Arbitrary trees can be created effectively and the blend of enormous units of irregular trees normally prompts right models. Random tree models have been broadly created in the field of AI to manufacture a reasonable and exact model for different characterizations

Random Forest

One of the ensemble techniques referred to as random forests envision that every one of the classifiers in the gathering is a choice tree classifier with the goal that the accumulation of classifier techniques is to Improve Classification Accuracy, is a "backwoods." The character decision shrubs have created the utilization of an arbitrary assurance of credits at each hub to choose the split. All the more officially, every tree depends upon the estimations of an irregular vector inspected freely and

with the equivalent conveyance for all timber in the woods. During order, each vote and most prominent class is returned.

Reduced Error Pruning Tree

Reduced Error Pruning (REP) Tree is the least complex and most understandable methodology in decision tree pruning. It is a quick decision tree student, which manufactures a choice or a relapse tree utilizing information procure as the splitting standard and prunes it the utilization of diminished blunder pruning. Utilizing REP calculation, the tree traversal has completed from posterior to zenith, and after that tests for each inside hub, and change it with frequent arrangements with most circumstance about the tree exactness and change it with regular characterization, with most circumstance about the tree precision, which should now lessen. The strategy will proceed till any likewise pruning will diminish exactness.

Naive Bayes Tree

A Naive Bayes (NB) classifier is a basic probabilistic classifier dependent on applying Bayes' hypothesis with autonomy suppositions study. Naive Bayes classifier output deals with a subjective assortment of unprejudiced factors, non-stop or straight out. The calculation makes forecasts utilizing Bayes Theorem which fuse proof or earlier information in its expectation.

Given a set of variables = {x₁,x₂,.....,x_d}, the posterior probability can be constructed for the event C_j amongst a set of possible consequences C = {c₁, c₂,....., c_d}. Simply put X is the predictor and C is the set of express stages presenting the established variable.

Utilizing Bayes rule: P(C_j/x₁,x₂,...x_d) αp(x₁,x₂,...x_d/C_j)/p(C_j),where p(C_j|x₁,x₂,.....,x_d) is the posterior probability of class participation.

Patient Dataset

The complete 400 cases with 25 special attributes was amassed from the kidney most cancer data set from kaggle. The attribute “diagnosis” is described as the measurable are with free sure imply person with kidney disease and no means that the person is no longer with kidney disease. Table 1 suggests the attributes/values of kidney disease dataset. The dataset having sure cases are 147 and 251 no cases.

Table 1: Kidney Dataset

| Serial Number | Attribute Name |
|---------------|------------------------|
| 1 | Age |
| 2 | blood pressure |
| 3 | specific gravity |
| 4 | Albumin |
| 5 | Sugar |
| 6 | red blood cells |
| 7 | pus cell |
| 8 | pus cell clumps |
| 9 | Bacteria |
| 10 | blood glucose random |
| 11 | blood urea |
| 12 | serum create nine |
| 13 | Sodium |
| 14 | Potassium |
| 15 | Haemoglobin |
| 16 | packed cell volume |
| 17 | white blood cell count |

| | |
|----|-------------------------|
| 18 | red blood cell count |
| 19 | Hypertension |
| 20 | diabetes mellitus |
| 21 | coronary artery disease |
| 22 | Appetite |
| 23 | pedal edema |
| 24 | Anemia |
| 25 | Class |

Proposed Technique

The main aim of this examination is to propose a technique that can create Classification Association Rules (CARs) productively and measure which strategy gives more level of right anticipated incentive for early conclusion of kidney malady. The relative investigation of the proposed strategy has been finished with other cutting edge methods. The concise subtleties of different advances are depicted as pursued:

Selection

There is a choice of information collection for expectation of kidney illness, to plan information examination and to get powerful learning. The adequate amount of information is required to perform information strategies to chose kidney dataset.

Pre-Processing and Transformation

The dataset is set up in ARFF (Attribute-Relation File Format) document position standard of kidney malady dataset. The information is changed over into right position for execution of cooperative methods. Different things required are the expulsion of right qualities for missing records, copy records, evacuate pointless information field, standard information position, adjust information in a convenient way and so on.

The execution of Bayes familiar computation and particularly wide range of Bayes estimation, which joins Naïve Bayes, Naïve Bayes Simple, Naïve Bayes Updatable and ADTree, Decision Stamp, FT,J₄₈,J₄₈ Graft, LAD Tree, LMT,NB Tree, Random Forest, Random Tree, REP Tree is done and need to pick the 10 best rules from each system for setting up the readiness instructive accumulation for use of different course of action techniques.

Selection of Associative Rules

The execution of Bayes acquainted calculation and distinctive variety of different Bayes calculation which incorporates Naïve Bayes, Naïve Bayes Simple, Naïve Bayes Updatable and ADTree, Decision Stamp, FT,J₄₈,J₄₈Graft, LAD Tree, LMT,NB Tree, Random Forest, Random Tree, REP Tree are done and need to choose the 10 best guidelines from every strategy for setting up the preparation for information collection for usage of various arrangement strategies.

Performance Evaluation

The grouping calculations like Naïve Bayes, Naïve Bayes Simple, Naïve Bayes Updatable and ADTree, Decision Stamp, FT, J₄₈,J₄₈ Graft, LAD Tree, LMT,NB Tree, Random Forest, Random Tree, REP Tree are actualized on preparing dataset and the yield of every calculation is assessed of the premise of remedied ordered occurrences.

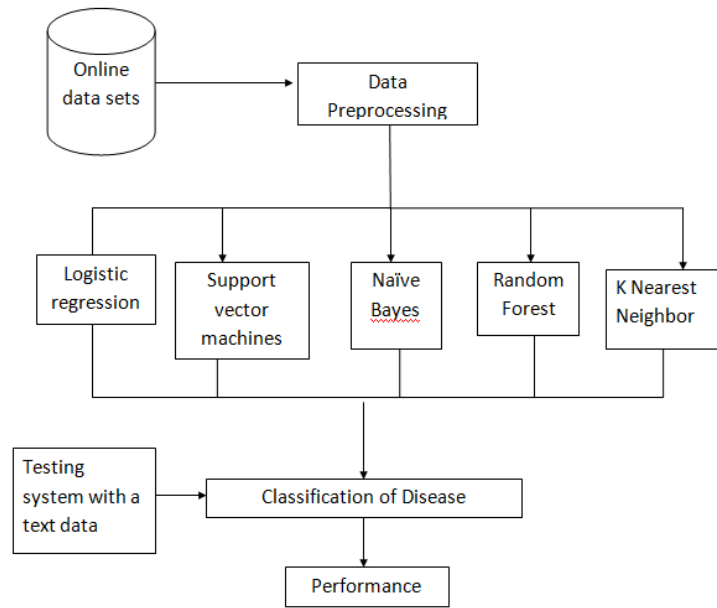


Figure 1: Performance of Machine Learning Algorithms for Chronic Kidney Disease.

Analysis of Various Algorithms

Analysis of various classifiers using Waikato Environment Knowledge Analysis tool is presented and dataset is considered from chronic kidney dataset from kaggle.

Table 2: Analysis of Various Algorithms

| Classifier | Correctly Classified Instances (%) | Incorrectly Classified Instances (%) | Kappa Statistic | Mean Absolute Error | Root Mean Square Error | Relative Absolute Error (%) | Root Relative Square Error (%) |
|-----------------------|------------------------------------|--------------------------------------|-----------------|---------------------|------------------------|-----------------------------|--------------------------------|
| Bayes Net | 80.40 | 19.59 | 0.5986 | 0.1929 | 0.4077 | 41.38 | 84.47 |
| Naïve Bayes | 78.39 | 21.60 | 0.5489 | 0.2111 | 0.4276 | 45.28 | 88.60 |
| Naïve Bayes Simple | 78.64 | 21.35 | 0.5535 | 0.2163 | 0.4341 | 46.41 | 89.93 |
| Naïve Bayes Updatable | 78.39 | 21.60 | 0.5489 | 0.2111 | 0.4276 | 45.28 | 88.60 |
| AD Tree | 81.40 | 18.59 | 0.6086 | 0.2578 | 0.3611 | 55.31 | 74.81 |
| Decision Stump | 79.14 | 20.85 | 0.5867 | 0.2814 | 0.3779 | 60.49 | 78.30 |
| FT | 81.15 | 18.84 | 0.5972 | 0.19995 | 0.3928 | 42.80 | 81.38 |
| J48 | 81.65 | 18.34 | 0.6144 | 0.2236 | 0.367 | 47.96 | 76.04 |
| J48 Graft | 81.15 | 18.84 | 0.605 | 0.222 | 0.3672 | 47.68 | 76.07 |
| LAD Tree | 77.88 | 22.11 | 0.5294 | 0.2568 | 0.3912 | 55.09 | 81.05 |
| LMT | 82.41 | 17.58 | 0.6267 | 0.2153 | 0.3383 | 46.20 | 70.09 |
| NB Tree | 76.88 | 23.11 | 0.5134 | 0.2595 | 0.4105 | 55.68 | 85.04 |
| Random Forest | 83.41 | 16.58 | 0.647 | 0.2372 | 0.3372 | 55.90 | 69.86 |
| Random Tree | 76.13 | 23.86 | 0.481 | 0.246 | 0.4412 | 52.77 | 91.91 |
| REP Tree | 80.42 | 19.59 | 0.5931 | 0.25 | 0.3741 | 53.64 | 77.50 |

Total number of instances is 398 and ignored class instances are two.

Table 3: Performance Measurement of various Algorithms

| | True Positive Rate | False Positive Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---------------------------|--------------------|---------------------|-----------|--------|-----------|----------|-------|
| <i>Bayes Net</i> | 0.85 | 0.223 | 0.691 | 0.85 | 0.762 | 0.895 | Yes |
| | 0.777 | 0.55 | 0.899 | 0.777 | 0.833 | 0.886 | No |
| Weighted Average | 0.804 | 0.177 | 0.822 | 0.804 | 0.807 | 0.889 | |
| <i>Naïve Bayes</i> | 0.776 | 0.211 | 0.683 | 0.776 | 0.726 | 0.884 | Yes |
| | 0.789 | 0.224 | 0.857 | 0.789 | 0.822 | 0.874 | No |
| Weighted Average | 0.784 | 0.22 | 0.793 | 0.786 | 0.786 | 0.878 | |
| <i>Naïve Bayes Simple</i> | 0.776 | 0.207 | 0.687 | 0.776 | 0.728 | 0.876 | Yes |
| | 0.743 | 0.224 | 0.858 | 0.793 | 0.824 | 0.867 | No |
| Weighted Average | 0.786 | 0.218 | 0.795 | 0.786 | 0.784 | 0.87 | |
| <i>Bayes Updatable</i> | 0.776 | 0.211 | 0.683 | 0.776 | 0.726 | 0.874 | Yes |
| | 0.789 | 0.244 | 0.857 | 0.789 | 0.822 | 0.879 | No |
| Weighted Average | 0.784 | 0.22 | 0.793 | 0.784 | 0.786 | 0.878 | |
| <i>AD Tree</i> | 0.796 | 0.715 | 0.727 | 0.796 | 0.76 | 0.892 | Yes |
| | 0.825 | 0.204 | 0.873 | 0.848 | 0.848 | 0.883 | No |
| Weighted Average | 0.814 | 0.193 | 0.819 | 0.816 | 0.816 | 0.886 | |
| <i>Decision Stump</i> | 0.918 | 0.283 | 0.655 | 0.918 | 0.765 | 0.824 | Yes |
| | 0.717 | 0.082 | 0.938 | 0.717 | 0.813 | 0.827 | No |
| Weighted Average | 0.789 | 0.156 | 0.833 | 0.791 | 0.791 | 0.819 | |
| <i>FT Tree</i> | 0.755 | 0.155 | 0.74 | 0.755 | 0.747 | 0.836 | Yes |
| | 0.845 | 0.245 | 0.855 | 0.845 | 0.85 | 0.8219 | No |
| Weighted Average | 0.812 | 0.212 | 0.812 | 0.812 | 0.812 | 0.831 | |
| <i>J48</i> | 0.803 | 0.175 | 0.728 | 0.803 | 0.764 | 0.888 | Yes |
| | 0.825 | 0.197 | 0.877 | 0.825 | 0.85 | 0.879 | No |
| Weighted Average | 0.817 | 0.189 | 0.822 | 0.817 | 0.818 | 0.882 | |
| <i>J48 Graft</i> | 0.803 | 0.183 | 0.72 | 0.803 | 0.759 | 0.889 | Yes |
| | 0.817 | 0.197 | 0.876 | 0.817 | 0.843 | 0.88 | No |
| Weighted Average | 0.812 | 0.192 | 0.818 | 0.812 | 0.813 | 0.883 | |
| <i>LAD Tree</i> | 0.721 | 0.187 | 0.693 | 0.721 | 0.707 | 0.866 | Yes |
| | 0.813 | 0.279 | 0.833 | 0.813 | 0.823 | 0.855 | |
| Weighted Average | 0.779 | 0.245 | 0.781 | 0.781 | 0.78 | 0.78 | |
| <i>LMT Tree</i> | 0.789 | 0.155 | 0.748 | 0.768 | 0.768 | 0.916 | Yes |
| | 0.845 | 0.211 | 0.872 | 0.858 | 0.858 | 0.908 | No |
| Weighted Average | 0.824 | 0.19 | 0.827 | 0.824 | 0.825 | 0.911 | |
| <i>NB tree</i> | 0.735 | 0.211 | 0.671 | 0.735 | 0.701 | 0.846 | Yes |
| | 0.789 | 0.265 | 0.835 | 0.789 | 0.811 | 0.838 | No |
| Weighted Average | 0.769 | 0.245 | 0.775 | 0.769 | 0.771 | 0.841 | |
| <i>Random Forest</i> | 0.786 | 0.143 | 0.765 | 0.796 | 0.78 | 0.918 | Yes |

Table 3 Contd.,

| | | | | | | | |
|--------------------|-------|-------|-------|-------|-------|-------|-----|
| | 0.857 | 0.204 | 0.878 | 0.857 | 0.867 | 0.908 | No |
| Weighted Average | 0.834 | 0.182 | 0.836 | 0.834 | 0.835 | 0.912 | |
| <i>Random Tree</i> | 0.646 | 0.171 | 0.688 | 0.646 | 0.667 | 0.822 | Yes |
| | 0.829 | 0.354 | 0.8 | 0.829 | 0.814 | 0.817 | No |
| Weighted Average | 0.761 | 0.286 | 0.759 | 0.761 | 0.76 | 0.819 | |
| <i>REP Tree</i> | 0.816 | 0.203 | 0.702 | 0.816 | 0.755 | 0.869 | Yes |
| | 0.797 | 0.184 | 0.881 | 0.797 | 0.837 | 0.86 | No |
| Weighted Average | 0.804 | 0.191 | 0.815 | 0.804 | 0.806 | 0.863 | |

Performance Measures for Classification

One can utilize the following execution measures for the grouping and forecast of issue-inclined module in agreement with his/her own special need. Confusion Matrix: The disarray framework is utilized to quantify the general execution of two kind of issues for the given informational collection. The correct slanting components TP (genuine positive) and TN (genuine negative) adequately arrange instances just as FP (false positive) and FN (false negative) erroneously characterize instances. Confusion Matrix Correctly Classify Instance TP+TN Incorrectly Classify Instances.

- True positives allude to the positive kidney tuples that were effectively named by the classifier,
- True negatives are the negative kidney tuples that were accurately named by the classifier.
- False positives are the negative kidney tuples that were inaccurately named as positive tuples
- False negatives are the positive kidney tuples that were erroneously marked negative tuples

A Confusion Matrix for Positive and Negative Tuples is as Follows

Table 4: Predicted Class Confusion Matrix

| | | Yes | No | |
|--------------|-----|----------------------|---------------------|-----|
| Actual Class | Yes | True Positives (TP) | False Negatives(FN) | P |
| | No | False Positives (FP) | True Negatives(TN) | N |
| | | P Complement | N Complement | P+N |

The table may have extra lines or segments to give sums. For instance, in the confusion matrix of above, Figures P and N appear. Also, P Complement is the quantity of tuples that were named as positive (TP+FP) and N Complement is the quantity of tuples that we relabeled as negative (TN+FN). The complete number of tuples is TP+TN+FP+FN, or P+N, or P Complement +N Complement. Note that in spite of the fact that the perplexity framework demonstrated is for a paired classification issue, confusion matrix can be effectively drawn for numerous classes along these lines. Presently, we should take a gander at the assessment measures, beginning with exactness. The precision of a classifier on a given test set is the level of test set tuples that are accurately classified by the classifier that is

Table 5: Various Performance Measurements

| Measure | Formula |
|-----------------------------------------|-----------------|
| Accuracy, Recognition Rate | $(TP+TN)/(P+N)$ |
| Error, Misclassification Rate | $(FP+FN)/(P+N)$ |
| Sensitivity, True Positive rate, Recall | TP/P |
| Specificity, True Negative Rate | TN/N |

Confusion Matrix all above algorithms are given below

Bayes Net

Table 6

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=125 | False Negative=22 | A=Yes |
| False Positive=56 | True Negative =195 | B=No |

Naïve Bayes

Table 7

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=114 | False Negative=33 | A=Yes |
| False Positive=53 | True Negative =198 | B=No |

Naïve Bayes Simple

Table 8

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=114 | False Negative=33 | A=Yes |
| False Positive=52 | True Negative =199 | B=No |

Naïve Bayes Updatable

Table 9

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=144 | False Negative=33 | A=Yes |
| False Positive=53 | True Negative =198 | B=No |

AD Tree

Table 10

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=117 | False Negative=30 | A=Yes |
| False Positive=40 | True Negative =207 | B=No |

Decision Stump

Table 11

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=135 | False Negative=12 | A=Yes |
| False Positive=71 | True Negative =180 | B=No |

Tree FT

Table 12

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=113 | False Negative=36 | A=Yes |
| False Positive=39 | True Negative =212 | B=No |

J48

Table 13

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=118 | False Negative=29 | A=Yes |
| False Positive=44 | True Negative =207 | B=No |

J48 Graft**Table 14**

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=118 | False Negative=29 | A=Yes |
| False Positive=46 | True Negative =205 | B=No |

LAD Tree**Table 15**

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=116 | False Negative=41 | A=Yes |
| False Positive=47 | True Negative =204 | B=No |

LMT Tree**Table 16**

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=116 | False Negative=31 | A=Yes |
| False Positive=39 | True Negative =212 | B=No |

NB Tree**Table 17**

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=108 | False Negative=39 | A=Yes |
| False Positive=53 | True Negative =198 | B=No |

Random Forest**Table 18**

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=117 | False Negative=30 | A=Yes |
| False Positive=36 | True Negative =215 | B=No |

Random Tree**Table 19**

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=95 | False Negative=52 | A=Yes |
| False Positive=43 | True Negative =208 | B=No |

REP Tree**Table 20**

| A=Yes | B=No | Classified as |
|-------------------|--------------------|---------------|
| True Positive=120 | False Negative=27 | A=Yes |
| False Positive=51 | True Negative =200 | B=No |

Correctly and Incorrectly Classified Instances

Correctly classified instances mean the sum of True Positives and True Negatives of kidney dataset tuples. Similarly, incorrectly classified instances means the sum of false positive and false negatives of kidney datasets. The total number of correct kidney data instances divided by total number of kidney data instances gives the accuracy.

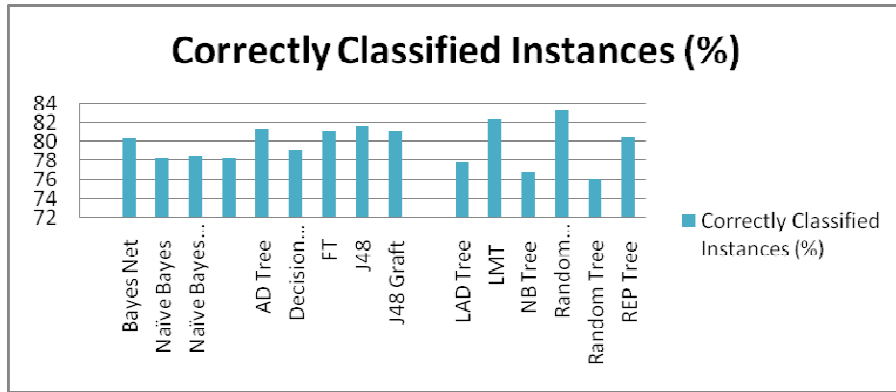


Figure 2: Comparison of Correctly Classified Instances.

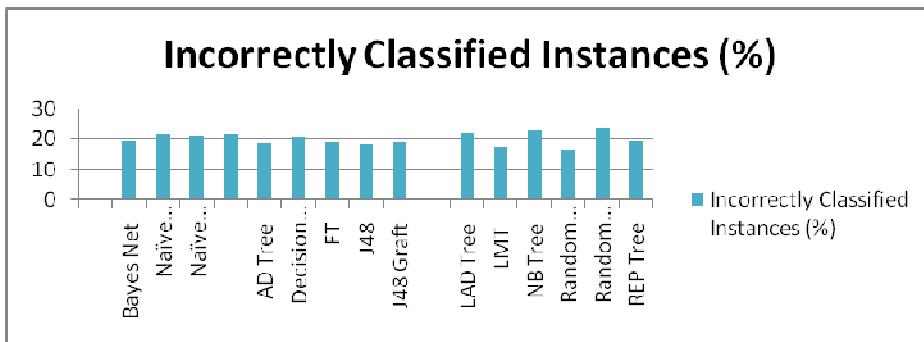


Figure 3: Comparison of Incorrectly Classified Instances.

Kappa Statistic

The **Kappa Measurement** is a proportion of how intently the kidney information occasions grouped by the machine learning classifier coordinated the kidney information named as ground truth, controlling for the precision of an irregular classifier as estimated by the normal exactness.

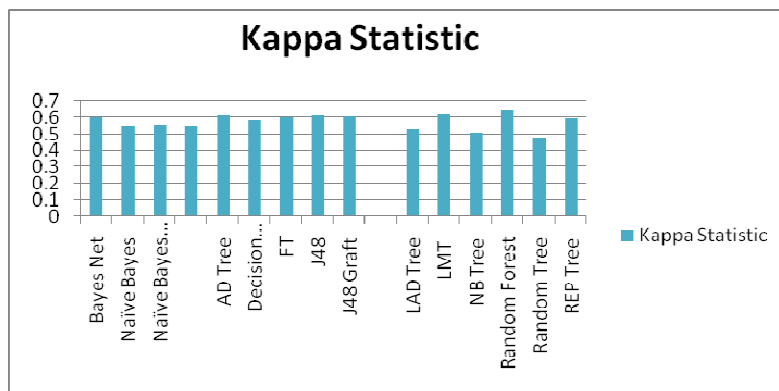


Figure 4: Comparison of Kappa Statistic.

Mean Absolute Error

Given the kidney test informational index, **Mean Absolute Error** of your model alludes to the **mean** of the supreme estimations of every forecast **error** on all occurrences of the kidney test informational collection

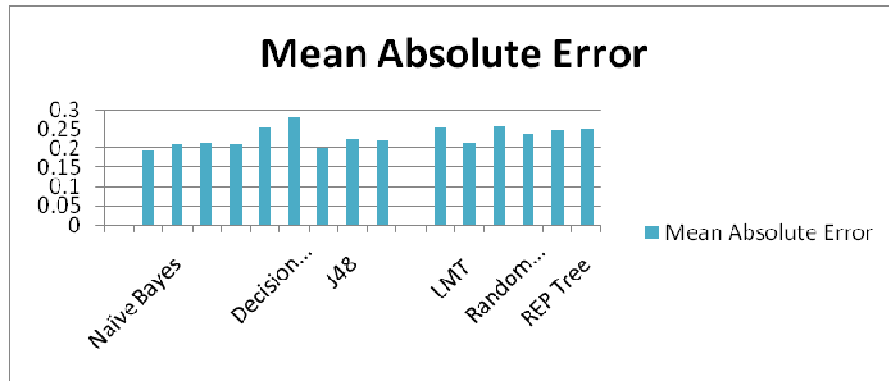


Figure 5: Root Mean Square Error.

Relative Absolute Error

Relative Absolute Error (RAE) is an approach to gauge the exhibition of a prescient model. **The Relative Absolute Error** is communicated as a proportion, looking at a mean error (leftover) to blunders created by an insignificant or gullible kidney informational index tuples

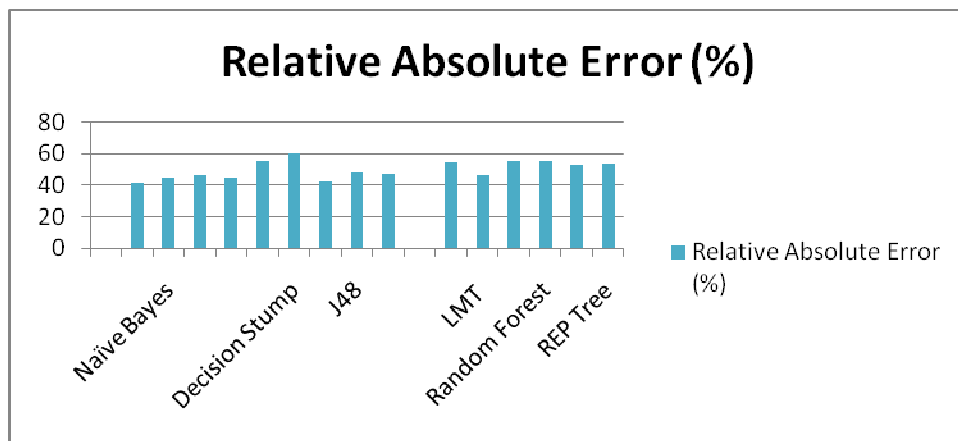


Figure 6: Comparison of Relative Absolute Error.

Root Mean Square Error

The root-mean-square error (RMSE) is an as often as possible utilized proportion of the contrasts between qualities anticipated by a model and the qualities really watched tuples of kidney dataset.

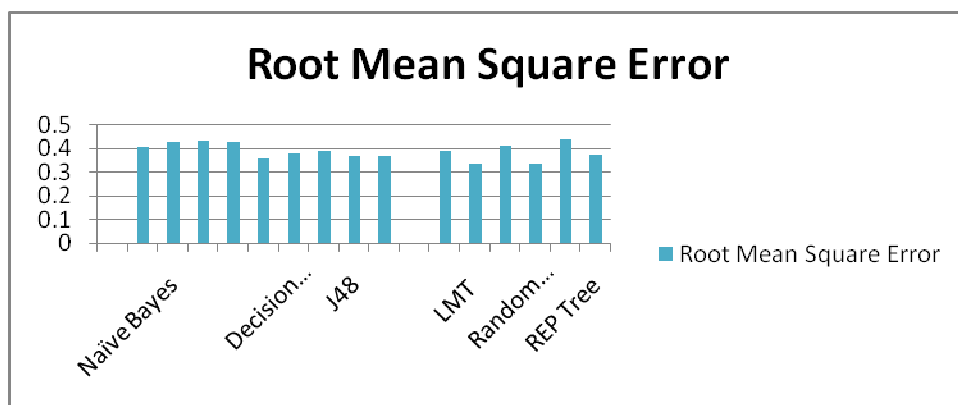


Figure 7: Comparison of Root Mean Square Error.

DISCUSSIONS

In this study, we have a tendency to apply machine learning algorithms for chronic kidney dataset to predict whether patients have chronic kidney disease, and people who do not seem to be sick, supported the information of every attribute for every patient. Our goal was to match totally different classification models and outline the foremost economical one. Our comparison was created on the premise of several algorithms, which include Naïve Bayes, Naïve Bayes Simple, Naïve Bayes Updatable and ADTree, Decision Stamp, FT,J48,J48 Graft, LAD Tree, LMT,NB Tree, Random Forest, Random Tree,REP Tree.

Regarding accuracy, that represents the proportion of instances classified properly, we have a tendency to notice a variation between 76 and 83. This has no relationship with the classifiers; however, it is with application domain and sort of knowledge. In our study, Random Forest scored a decent accuracy (83.41%) followed by LMT (82.41%), ADTree (81.40%), J48 (81.15%), J48 graft (83.15%), Bayes (80.40%) are higher than 80%.

With respect to rate, Random Forest denoted the lowest error rate (16.58%) and accordingly the highest one was scored by Random Tree (23.86%). The letter of information point worth demonstrates that the value of all indicators is higher than 0.50 aside from Random Tree (0.481). This infers our classifiers are brilliant per degree scale anticipated via (Landis and Koch) [24], then again, actually Random Forest scored the best expectation understanding and identifying with the proportion of indicators, the estimations of mean absolute error (MAE), Root Mean Square Error (RMSE), Relative Outright Error (RAE), Root Relative Square Error (RRSE) demonstrated that C4.5 indicators scored unsurpassed low qualities (MAE = 0.2372) (RMSE = 0.3372, RAE = 55.90%,79, RRSE = 91.4185%) trailed by LMT,AD Tree,J48,J48 Graft.

Another necessary live are F-Measures which mix two performance measures: preciseness and recall. If we have a tendency to take the case of expected patients with the unwellness Random Forest marked the most effective rate (0.78), and within the case of non-disease, it marked the most effective rate additionally (0.867).

The confusion matrix demonstrates that all the calculations are grouped (398) examples appropriately with a couple of misclassified cases. Irregular forest is subtracted on the grounds that the best and as far as the most noteworthy assortment of occasions are appropriately arranged and in this way, there is least blunder rate at the expectation. It furthermore is the essential one in exactness and has the best f-measures rate, with an OK rate time of execution. LMT is hierarchal in light of the fact that the second once Random Forest, anyway outflanks in structure time of the arrangement and precision. Arbitrary Forest has demonstrated its exhibition as a solid classifier in terms of exactness and in this manner, the base execution time, which makes it a respectable classifier to be utilized in the therapeutic field for order and forecast

CONCLUSIONS

As conclusion, the learning machine digging strategies for prophetic investigation is unbelievably fundamental inside the wellbeing field because this offers us the capacity to confront sicknesses prior, thus spare individuals' lives through the expectation of fix. During this work, we tend to utilize many learning rules, Random Forest, LMT, FT, J48, J48 Graft, NB, to anticipate patients with constant kidney infection, and patients do not appear to be stricken by this unwellness. Reproduction results demonstrated that Random Forest classifier demonstrated its exhibition in anticipating the best prompts in terms of precision and least execution time.

REFERENCES

1. John, R., Webb, M. and Young, A., "Un referred chronic urinary organ disease: a longitudinal study", *AJU*, Vol. 5, Issue no. 3, 2004, pp. 825–35.
2. Sinha, Parul, and Poonam Sinha. "Comparative study of chronic uropathy prediction exploitation KNN and SVM." *IJEAT*, Vol. 4, Issue no. 12, 2015, pp. 608–12.
3. S. Ramya and Dr. N. Radha, "Diagnosis of Chronic Uropathy Exploitation Machine Learning Algorithms", *Int. J. Innov. Anal. Laptop Commun. Eng.*, Vol. 4, Issue no. 1, 2016, pp. 812–820.
4. Agrawal, V., Jaar, B. G., Frisby, X. Y., et al. Access to health care among adults evaluated for CKD: findings from the urinary organ – Early Analysis Program (KEEP). *Am. J. Urin. Organ Dis.* 2012; Vol. 59, Issue 3, pp. S5-S15.
5. Moumita Bhattacharya¹, Claudine Jurkovitz, M.D., and Hagit Shatkay, Assessing Chronic uropathy from workplace visit records exploitation class-conscious. *Meta-Classification of associate unbalanced Dataset*.
6. K. R. Ananthapadmanaban and G. Parthiban. Applying Machine Learning Techniques for Predicting the chance of Chronic uropathy, *Ind. J. Sci. Technol.*, Vol. 9, Issue no. 29, August 2016, DOI: 10.17485/ijst/2016/v9i29/93880,
7. Jyoti Saini¹, R. C. Gangwar² and Mohit Marwaha³, A Unique Detection for Pathology Exploitation Improved Support Vector Machine, *IJLTET*, Vol. 8, Issue no. 1, pp. 114–121.
8. Sirage Zeynu and Shruti Patil, Prediction of Chronic uropathy exploitation data processing Feature choice and Ensemble methodology, *WSEAS TRANSACTIONS on informatics and APPLICATIONS*, Vol. 15, 2018, E-ISSN: 2224–3402.
9. Pratibha Devishri, S., Ragin, O. R., and Anisha, G. S., Comparative Study of Classification Algorithms in Chronic uropathy, *IJRTE*, Vol. 8, Issue no. 1, May 2019, ISSN: 2277–3878.
10. Merve Doğruyol Başar¹ and Aydın Akan, Chronic uropathy Prediction with Reduced Individual Classifiers, analysis Article, *Electrica* Vol. 18, Issue no. 2, 2018, pp. 249–255.

AUTHOR PROFILE



Dr. M. Srivenkatesh working as Associate Professor, Department of Computer Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India .He has published nine international Journal papers. His research interest includes Data mining, Machine Learning, and Software Engineering, Cloud Computing and Rough Sets. Seven Research Scholars are working for their Ph.D. under his guidance.

