INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

A Web Application for Predicting Chronic Kidney Diseases Using Machine Learning

Lavanya M[1], Vinay Prasad M S[2]

[1]M Tech Student [2]Assistant Professor, Department of Electronics and Communication, JSS Science and Technology University Mysore, India.

Abstract: Chronic Kidney Disease(CKD) has turn into a serious issue with high growth rate, which requires kidney transplant and Dialysis. So early detection is required to overcome CKD. In this Project, we proposed two ML algorithms to predict CKD, one of the algorithm i.e, Naive Bayes yields a results with an accuracy of 91.54 percentage. comparing with same, KNN yields a better result with an accuracy of 97.18 percentage. GFR (glomerular filtration rate) is used to detect CKD in its early stage by measuring kidney functional level. Diet is recommended based on the stage. we developed a web UI for doctor who can measure kidney function level, stages using GFR equation, upload patient treatment and suggest a diet plan for the patient.
Keywords: CKD, GFR, KNN, Naive Bayes.

I. INTRODUCTION

Chronic Kidney disease is a major area of interest since its an global health issue. Among one in ten people have CKD and one in three people is at increased risk of CKD, Worldwide highest rate of CKD is about 24 percentage found in sauid Arabia and Belgium [8]. As reported by the GBDS(Global Burden of Disease Study) in 2010, from 27th in 1990 to 2010,CKD is determined as major cause for mortality [5].Chronic kidney disease affected people worldwide is around 500 million people[5]. Chronic Kidney disease is rising from year to year in developing country like Bangladesh, where the total population of CKD is determined as 14 percent- age including Bangladesh[5].Another Study showed that 26 percentage of chronicity is discovered in 30 year old and above people in urban Dhaka region and 13 percentage chronicity in 15 year and above people. In 2013, chronicity study in Bangladesh exposed that risk rate of residents in rural is around one-third due to misdiagnosed CKD at the time. Due to Untreated or failure of Kidney, around one million people die worldwide, Good Quality of life can be achieved by early detection of CKD i.e, stage prediction which can be treated by lowering the blood pressure, drugs, diet and lifestyle changes. if patient with CKD is not treated they have 20 times more chance to die, so early diagnosis and treatment is important. Different Classification approaches for data mining and machine learning algorithms are used to detect chronic disease. chronic disease is nothing but a disease that last over a long period time which requires proper medical attention and lifestyle change or both may be requires in some cases. This may result into a death and disability. In our work we are concentrating more on chronic renal disease which is also known as chronic kidney disease. current medical system requires more time for CKD prediction as it requires more time for diagnosis, consultation and experience etc. Identification of Chronic Kidney disease in early stages is challenging factor in the current medical sector. CKD is a condition where Kidneys cannot function properly i.e, it cannot filter toxic from the body. Our Work predominantly focuses on detecting life threatening disease like Chronic renal Disease using machine learning algorithms and predicts the stage of CKD using GFR which helps in early detection and reduces the severity of CKD. Diet Management of CKD Patient depends on GFR rate and Disease Severeness of CKD is classified into 5 stages. where stage 1 is safe, stage 2 is moderate which requires strict diet plan, stage 3 stage 4 and stage 5 are difficult for patients to balance electrolyte, minerals and liquids[11], GFR can be calculated using formula[9]. Our proposed system is a browser based application which can be accessed by doctors from different location at anytime.

II. LITERATURE SURVEY

In this section we like to provide existing and current technologies used for prediction of CKD ,[1] to predict kidney disease they used BP neural network and Logistic Regression model based on the dataset. They analyzed the relationship between various attributes in the dataset and the status of patient disease. They found that BP neural network model achieved better accuracy, error rate and recall. By using BP neural network CKD can be detected easily by the medical practitioner in a short period of time with more accuracy, so patient can be treated more timely, For the proposed system an hybridized ML algorithms can be used as future scope.

The study [2] shows that in developing countries J48 decision tree is more suitable Machine Learning Algorithm, due to quick application of the algorithm. J48 yields an accuracy of 95 percentage which is agreed by the experienced nephrologist. Inversely, Random Forest, Naive Bayes, Support Vector Machine, K-nearest neighbor and Multilayer perceptron yields an accuracy of 93.33, 88.33, 76.66, 71.67 and 75 percentage respectively. Limitation of the proposed model is that the dataset is relatively small for evaluation. The proposed work [3] presents an system which can analyze seven different machine learning techniques which includes Naive Bayes J48 Decision Tree, Multilayer Perception, Logistic Regression, support vector Machine Naive Bayer Tree and Composite Hypercube or Integrated Random Projection. As an evaluation metrics, Root Mean squared error, Root Relative Squared Error, (PNN)Mean Absolute Error, Relative Squared Error, Accuracy, Recall, F-measures and Precision. The experimental result shows that CHIRP decrease error rate and enhance the accuracy for detecting CKD.[4] aim is to identify the best suited classification algorithm for diagnosis of CKD based upon reports of classification and factors of performance, related work is performed on different classification algorithms like Random Forest, XG Boost, Logistic Regression, Naive Bayes, Neural networks and Support Vector Machine classifier. From the experimental results Random Forest and XG Boost achieves a better result of 99.29 percentage accuracy compared to other classification algorithms. [5] they proposed a system which can detect CKD at mild stage. More accurate machine learning methods and Data mining are needed to diagnose and predict CKD. The model is evaluated by using 7-fold and 10- fold cross validation procedure. Different ML algorithms are used in the model which includes K-Nearest Neighbor (K- NN), Discriminant Analysis(DA), Support Vector Machine, Naive Bayes and Random Forest. This experiment is con- ducted in MATLAB tool, The result shows that Random Forest achieves better results. They proposed [6] a system which can predict CKD, they used datasets from UCI Machine Learning Repository and some real-time data from medical college. They used python for developing a system. Data is trained by using 10-fold CV and it is applied on Random Forest and ANN classification algorithms. Hybridized algorithms can be used for the proposed system as a future work.[7] performance is evaluated based on bagging ensemble technique, which is applied on CKD dataset for feature selection. To improve the learners performance K-Nearest Neighbor, Naive Bayes and Decision Tree algorithms, performance were aggregated by sing bagging ensemble approach. As a result of this model achieves 100 percentage accuracy for CKD prediction compared with 98.3 percentage without feature selection.[8] various factors responsible for CKD, so prediction of CKD in early stage is much needed.

For evaluation of the model they used Naive Bayes, KNN and Random Forest classification algorithms, among this algorithms Random Forest performs better. From this model CKD can be treated in timely manner before disease getting worse.[9] filling of missing values of CKD, similar measurements can be used to process the missing data. Once the missing dataset is filled effectively, six machine learning algorithms are applied to the model. Among six classifier algorithms Random Forest yields better result of 99.75 percentage. Hybridized ML model is proposed by combining logistic regression and Random Forest which achieves an accuracy of 99.83 percentage after simulating it for 10times which is a major drawback of the proposed system. [10] the experimental result shows that the accuracy achieved by probabilistic Neural Networks is more when compared with other classifiers to predict CKD on the other side the execution time required by probabilistic Neural Network is about 12s which is more when compared with multilayer perceptron, multilayer perceptron requires 3s for its execution. The PNN algorithm achieves a better accuracy and performance in predicting CKD. For future work their execution time can be reduced. System [11] includes 4 modules they are data preprocessing, feature extraction, defining stages based on potassium level, diet recommendation respectively. Dataset includes 25 features appropriate parameter and highly effective statistical method like regression extraction is used for CKD detection. Main goal of machine learning module is to detect the patient's status of CKD and classify them into various stages. Diet recommendation is done purely on blood potassium level which is the major drawback of the proposed system. Aim [12]is to detect chronic kidney disease by using some of the Machine Learning Algorithms and feature selection methods. The objective is to combine different features and input them into ML algorithms. The implementation of the algorithm is based on feature selection and then their performance is compared. For the proposed system, Decision tree, Random forest and Logistic regression yields an accuracy of 98.48, 94.16 and 99.24 percentage respectively. Precision of the classification algorithms are 100, 95.12 and 98.82, Recall is 97.61, 96.12 and 100 percentage respectively. Two techniques of feature selection are combined to strengthen the algorithm. Effectiveness of model can be increased by taking more datasets.

III. PROPOSED SYSTEM

The Proposed System is an mechanism for predicting Chronic Kidney Disease using different Classification Techniques. CKD can be Classified according to its severeness using Machine learning process.

Aim of this project is to predict CKD using different Machine learning algorithms, medical test records of CKD patient can be utilized to recommend dietplan by using different classification algorithm.

System uses old data from "UCI Repository" and uses tools such as "Visual Studio" and "SQL Server" to develop application. System is an real time application useful for doctors to identify CKD and related stages and recommending the suitable diet for the patients.

A. K-Nearest Neighbors

The k-nearest neighbors algorithm, also known as KNN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

Algorithm:

1) Select K-number of nearest neighbors.
2) Euclidean distance is calculated for selected K numberof nearest neighbors.
3) calculated Euclidean distance is used to get the K -nearest neighbors.
4) Number of data in every category is counted amongst K-nearest neighbors.
5) New data points are assigned to Maximum number of neighbor category.

B. Naive Bayes

Naive Bayes Classifier Algorithm is based upon Bayes theorem and is used to solve Classification Problem. Naive Bayes is an Probability Classifier, it is estimated based on the probability of the object.

Algorithm:

1) Data set are scanned from the storage server, data are retrieved from the server for mining purpose.
2) Each attribute value probability is calculated. 3) Apply the formula.

Bayes' theorem:

$P(A|B) = P(B|A) \cdot P(A) P(B)$ (1) where, $P(A—B)$: It's the probability of hypothesis A on the noticed B event. –

$P(B—A)$: It's the given data probability that the hypothesis probability is true.

$P(A)$: It's the hypothesis probability before noticing the data.

$P(B)$: proof of the data probability.

C. GFR

GFR can be estimated by using Modification of diet in Renal Disease and the Chronic Kidney Disease Epidemiology Collaboration equation (CKD-EPI). This are the most commonly used equations to evaluate the stages in CKD for patients of age 18 and above.

$GFR = 166 * (Scr/0.7)^{-0.329} * (0.993)^{Age}$ (2) Equation 2 is used when Serum Creatinine value is less than or equal to 0.7.

$GFR = 166 * (Scr/0.7)^{-1.209} * (0.993)^{Age}$ (3) Equation 3 is used when Serum Creatinine value is greater than 0.7.

$GFR = 163 * (Scr/0.9)^{-0.411} * (0.993)^{Age}$ (4) Equation 4 is used when Serum Creatinine value is less than or equal to 0.9.

$GFR = 163 * (Scr/0.7)^{-1.209} * (0.993)^{Age}$ (5) Equation 5 is used when Serum Creatinine value is greater than 0.9.

Missing values were managed by data pre-processing in the proposed method Binning Method is used Data binning, also called discrete binning or bucketing, is a data pre-processing technique used to reduce the effects of minor observation errors. It is a form of quantization.

The original data values are divided into small intervals known as bins, and then they are replaced by a general value calculated for that bin.

This has a soothing effect on the input data and may also reduce the chances of over fitting in the case of small data sets.90 percentage of data from the data sets are considered as training data sets 10 percentage as testing in the proposed system.

1) Patient data sets and parameters: In the first step of prediction process where we collect medical data. datasets were used for processing. Training data-sets will contains patient details and also parameters that are required for prediction which is shown in Table 1 and Table 2, where table 1 describes the attributes and their description used in the CKD prediction data sets. The above table 2 describes the attributes and their measurements used in the CKD prediction data sets.
2) Extract and Segment Data (Data Pre-processing): The medical data is analyzed and only relevant dataset are extracted. The data required for processing is extracted and segmented according to the requirement. This is done because entire training data not required for processing and if we input all data, it requires too much of time for processing, so data processing is done.

3) Train Data: Once required data extracted and segmented, we need to train the model, train means converting the data into the required format such as numerical values or binary or string etc. conversion depends on the algorithm type.

4) Classification Techniques: Supervised Learning Technique uses predictive method which includes the prediction of one value using other values from the dataset. Its object is to classify parameters based on predefined set of labels. Supervised learning can be built for multiple algorithms such as SVM, KNN, ID3, Regression techniques and Random Forest. Appropriate algorithms can be used for prediction of CKD depending on labels, parameters, dataset and requirement.
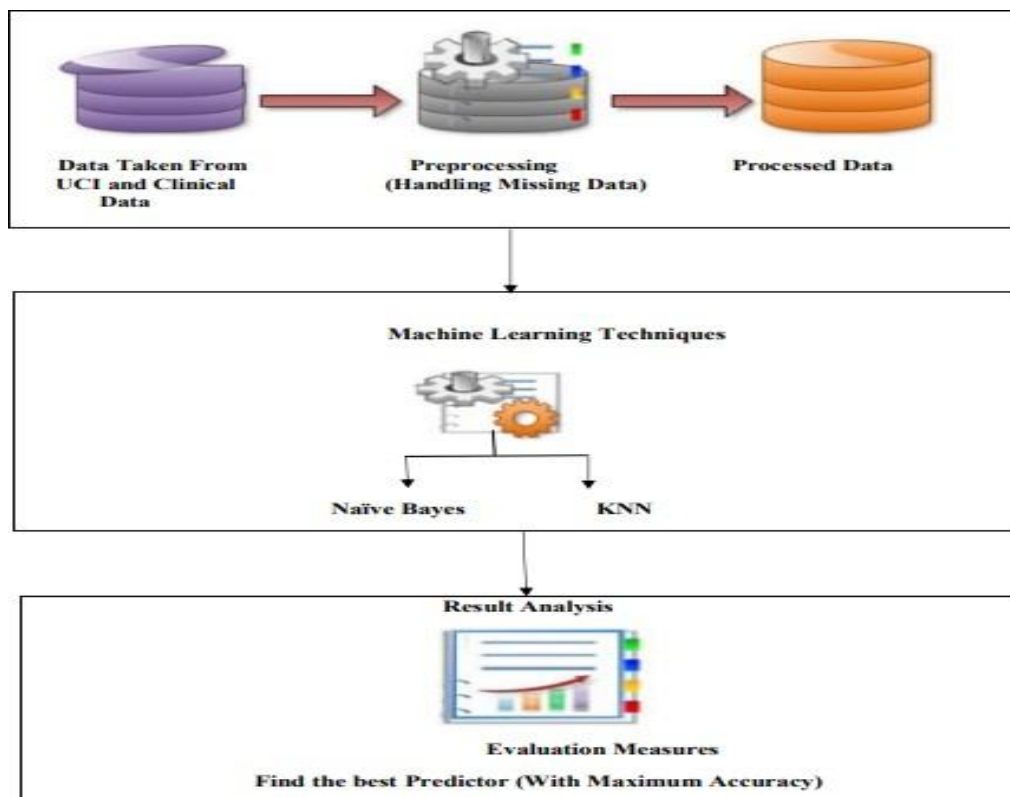
Fig. 1: Proposed Block Diagram.

5) The outcome of the proposed system is examined by using accuracy, precision, recall, F-measure and stage prediction is done by using GFR. Proposed system uses "KNN" algorithm or "Naive Bayes" for Disease prediction. the algorithm is checked with the accuracy using confusion matrix method. Here we validate the results generated by the algorithm "bayesian classifier" and "KNN algorithm".

TABLE I: Attributes and Description of CKD dataset.

Attribute	Description
age	Age
bp	Blood Pressure
sg	Specific Gravity
al	Albumin
rbc	Red Blood Cells
pc	Pus Cell
pcc	Pus Cell Clumps
ba	Bacteria
bgr	Blood Glucose Random
bu	Blood Urea
sc	Serum Creatinine
sod	Sodium

pot	Potassium
hemo	Hemoglobin
pcv	Packed Cell Volume
wc	White Blood Cell Count
rc	Red Blood Cell Count
htn	Hypertension
dm	Diabetes Mellitus
cad	Coronary Artery Disease
appet	Appetite
pe	Pedal Edema
ane	Anemia
class	Class
su	Sugar

TABLEII: Value range and Measurements of Attributes.

Attribute	Measurement and Value Range
age	Age in Years
bp	bp in mm/Hg
sg	1.005,1.010,1.015,1.020,1.025
al	0,1,2,3,4,5
rbc	normal, abnormal
pc	normal, abnormal
pcc	Present, notpresent
ba	Present, notpresent
bgr	mgs/dl
bu	mgs/dl
sc	mgs/dl
sod	mEq/L
pot	mEq/L
hemo	gms
pcv	numerical values
wc	cells/cumm
rc	millions/cmm
htn	yes, no
dm	yes, no
cad	yes, no
appet	good, poor
pe	yes, no
ane	yes, no
class	ckd, notckd
su	0,1,2,3,4,5

Chronic Kidney Disease (CKD) data set for analyzing experiment is obtained from UCI machine learning repository and clinical data set, Among 700 data sets 420 are diagnosed as having CDK and 280 as NCKD(no CKD),data set includes 25 attributes and 2 outcomes i.e., 'CKD' and ' NCKD'.

TABLE III: Confusion matrix terminology

	Class/Positive 1	Class/Negative0
Class1/Positive(1)	True Positive	False Positive
Class2/Negative(0)	False Negative	True Negative
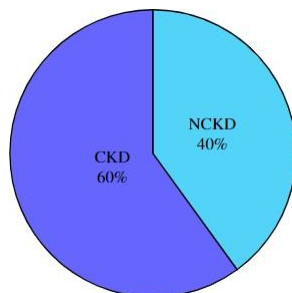
Fig. 2: Percentage of data belongs to CKD and NCKDclassification in a dataset

D. Users and Modules

1) *Admin:* Entire application is maintained by the Administrator. Administrator is an holder of the application.
2) *Doctor:* Doctor provides input for prediction of chronic renal disease. the major service provided by the system is prediction of chronic renal disease or CKD based upon the medical data.
3) *Receptionist:* Registration, billing and treatment details of patients are maintained by receptionist.
4) *Patient:* Patient is a one who uses application services.

E. Modules Of The Project Admin Has The Following Modules

1) *Login Module:* By inputting id and password admin can login into the application.
2) *Add Receptionists and Doctors:* n Number of doctors and receptionists can be added, Admin will be having authority to add receptionists and doctors.
3) *Set Id and Password of Staffs:* unique id and password for doctors and receptionist can be set by the admin.
4) *Add Parameters:* admin uploads the required parameters for CKD prediction, there are totally 24 parameters used for CKD prediction
5) *Change Password:* here admin can change the password
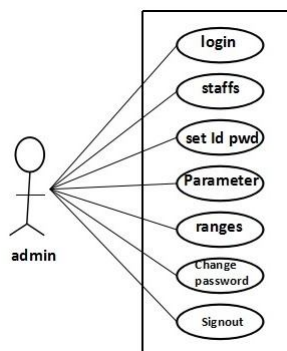
Fig. 3: Admin module.

F. Receptionist Has The Following Modules

1) *Login Module:* here receptionists get login to the application by inputting id and password.
2) *Data-set Module (Old Patients and parameters):* here receptionists upload the data set required for CKD prediction and stage prediction.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VIII August 2022- Available at www.ijraset.com

3) *Patient Registration:* Registration for new patient can be done here.
4) *Manage History of Patients*: Details of disease and treatment can be viewed in the patient history.
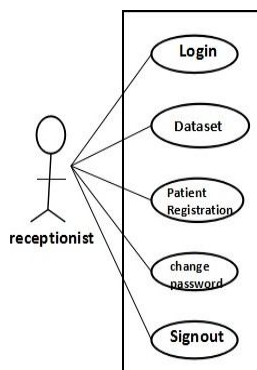5) *Change Password:* here admin can change the password.

Fig. 4: Receptionist module

G. *Doctor Has The Following Modules*
1) *Login Module:* By inputting unique id and password doctor can login into the application.
2) *Upload new Patient Data*: here parameters required for CKD prediction are inputted by the doctor.
3) *Chronic Kidney Disease Prediction Module [New patient – Naive Bayes Algorithm]:* this the core module where system predicts the CKD for the new patients based on the inputted parameters. Here system uses "Naive Bayes" algorithm for CKD prediction.
4) *Stage Prediction Module (GFR Method used):* in this module system predicts the stages for the patients suffering from CKD, system uses GFR method for stage prediction.
5) Diet Recommendation
6) *Treatment Details Upload:* treatment details of the patients are uploaded by the doctor.
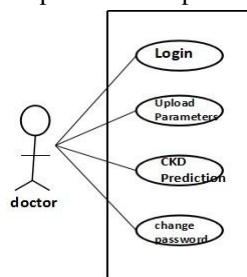
Fig. 5: Doctor module.

H. *Patient Has The Following Modules*
1) *Login Module*: here patient gets login to the application by inputting id and password.
2) *Patient History*: here patients can view the history such as registration details, treatment details, disease details etc.
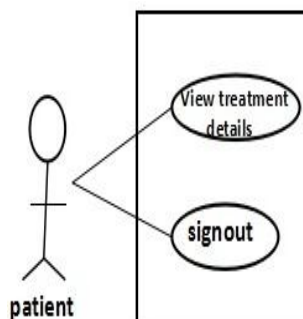3) *Treatment Module:* here patients can view the treatment details upload by doctor.

Fig. 6: patient module.

IV.   RESULTS AND DISCUSSION

This project uses CKD parameters for predicting CKD parameters for predicting CKD disease which helps medical students. this project is an automation for chronic renal disease prediction and disease complication, type can be identified from the database in an affective and economically fast¬ tracked manner, which can be achieved by applying naive bayes algorithm and KNN. Input to the algorithm is CKD parameters and disease is predicted based upon old CKD patient data.

A. Criteria For Assessment

Accuracy of the proposed technology can be measured by employing certain parameters like Precision, Recall, F-Measure and Accuracy. All this parameter can be measured by using the condition given below:

1) *Accuracy:* Accuracy is at the top priority in measuring performance and its a ratio perception that are correctly classified to the overall perception

   $Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$ (6)

   Equation 6 gives Accuracy, where total number of True-positive Classification is Given by TP, False-negative Classification is given by FN, True-negative Classification is given by TN, and False-positive Classification is given by FP.

2) *Precision:* Precision is the ratio of effectively expected observation that are positive to the overall positive observation.

   $Precision = \frac{TP}{TP + FP}$ (7)

   Equation 7 is used to find precision, where total number of True-positive Classification is Given by TP and False-positive Classification is given by FP.

3) *Recall:* It is a ration precisely expected observation that are positive to the observation in authentic class yes.

   $Recall = \frac{TP}{TP + FN}$ (8)

   Equation 8 is used to find recall, where total number of True-positive Classification is Given by TP and False-positive Classification is given by FP.

4) *F-Measure:* F-Measure is the average of Precision and Recall, it consider both false negative and false positives for its calculation. F-measure is more beneficial than accuracy, especially in the event where their is uneven distribution of class.

   $F - Measure = 2 (Recall * Precision) / (Recall + Precision)$ (9)

   Equation 9 is used to find F-measure.

TABLE IV: Assessment of Precision, Recall, F-Measure,Accuracy for Naive Bayes and KNN   classifier

Technique	Precision	Recall	F-Measure	Accuracy %
Naive Bayes	0.915	1	0.955	91.54
KNN	0.971	1	0.985	97.18

TABLE V: Assessment of different classifiers for instancewhich are correctly and incorrectly classified

Technique	Correctly Classified	Incorrectly Classified
KNN	679	21
Naive Bayes	641	59
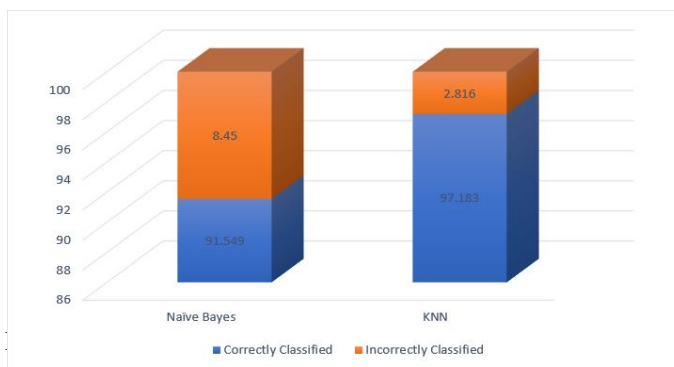
Constraint	Naive Bayes
Accuracy	91.5492957746479 %
Time (milli secs)	1911
Precision	0.915492957746479
Recall	1
F-Measure	0.955882352941176

KNN	Constraint
Accuracy	97.1830985915493 %
Time (milli secs)	2835
Precision	0.971830985915493
Recall	1
FMeasure	0.985714285714286
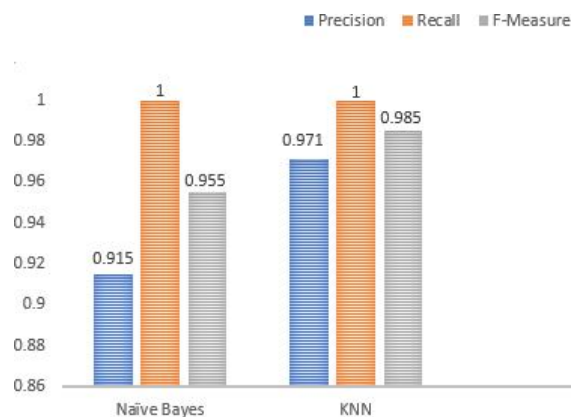
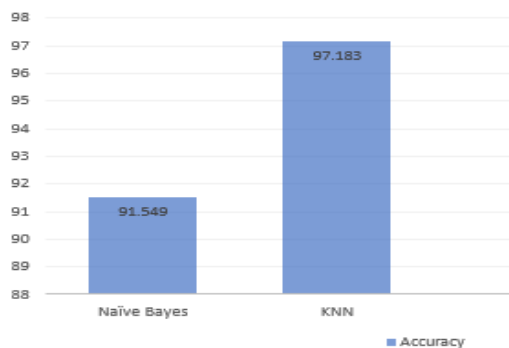Fig. 10: Analysis of Precision, Recall and F-measure

Fig. 11: Analysis of accuracy of Naive Bayes and KNN

Fig. 12: Stage prediction using GFR

Fig. 13: UI of Home Page

Fig. 14: UI of Doctor

V. CONCLUSION AND FUTURE WORK

In this paper, we employed a web application which can predict CKD, a comparative analysis was performed between KNN and Naive Bayes algorithm, in which KNN yield a better result. From the result obtained it is observed that overall precision, F-measure, recall is 0.971, 0.985, 1 and accuracy is 97.18 percentage respectively. this shows that KNN yields better result when compared with Naive Bayes. so medical practitioner can use KNN for prediction of CKD stage prediction is done by using GFR, depending on stage diet is recommended by the doctor and they can also upload patient's treatment details. patients can access uploaded treatment and recommended diet by the doctor. The proposed model uses a single algorithm to predict CKD, accuracy of the present work can be enhanced by using hybrid machine learning algorithms, which gives better accuracy compared with KNN algorithm. Along with this additional feature could be added to enhance our web application in future.

REFERENCES

[1] Bo Wang, "Kidney Disease Diagnosis Based on Machine Learning", 2020 International Conference on Computing and Data Science (CDS).

[2] ALVARO SOBRINHO , ANDRESSA C. M. DA S. QUEIROZ , LEANDRO DIAS DA SILVA, EVANDRO DE BARROS COSTA, MARIA ELIETE PINHEIRO ,AND ANGELO PERKUSICH, "Computer-Aided Diagnosis of Chronic Kidney Disease in Developing Countries: A Comparative Analysis of Machine Learning Techniques".

[3] BILAL KHAN,RASHID NASEEM,FAZAL MUHAMMAD,GHULAM ABBAS, AND SUNGHWAN KIM, "An Empirical Evaluation of Ma┐chine Learning Techniques for Chronic Kidney Disease Prophecy"

[4] N V Ganapathi Raju, K Prasanna Lakshmi, K. Gayathri Praharshitha, Chittampalli Likhitha,"Prediction of chronic kidney disease (CKD) using Data Science", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019).

[5] Inayatullah, Huma Qayyum, "An improved comparative model for chronic kidney disease (CKD) prediction" 2020, 14th International Conference on Open Source Systems and Technologies (ICOSST)

[6] Shanila Yunus Yashfi,Md Ashikul Islam,Pritilata,Nazmus Sakib ,Nazmus Sakib,Mohammad Shahbaaz ,Sadaf Salman Pantho,"Risk Prediction Of Chronic Kidney Disease Using Machine Learning Algorithms".

[7] Olayinka Ayodele Jongbo,Toluwase Ayobami Olowookere,Adebayo Olusola Adetunmbi, "Performance Evaluation of an Ensemble Method for Diagnosis of Chronic Kidney Disease with Feature Selection Technique",2020 International Conference on Decision Aid Sciences and Application (DASA).

[8] Devika R, Sai Vaishnavi Avilala, V. Subramaniyaswamy, "Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019).

[9] JIONGMING QIN, LIN CHEN, YUHUA LIU, CHUANJUN LIU, CHANGHAO FENG, BIN CHEN, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease".

[10] El-Houssainy A. Rady, Ayman S. Anwar,"Prediction of kidney disease stages using data mining algorithms".

[11] Akash Maurya,Rahul Wable,Rasika Shinde ,Sebin John ,Rahul Jadhav, Dakshayani.R, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using Machine Learning",2019 International Conference on Nascent Technologies in Engineering (ICNTE 2019).

[12] Rahul Gupta, Nidhi Koli, Niharika Mahor, N Tejashri ,"Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease",International Conference for Emerging Technology (INCET) .

[13] https://www.niddk.nih.gov/health-information/professionals/clinical-tools-patient-management/kidney-disease/laboratory-evaluation/glomerular-filtr

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ◎ (24*7 Support on Whatsapp)