International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-03, JUNE 2021

Chronic Kidney Disease Prediction Using Machine Learning

Gayatri Nayak, Student, Mumbai, India, nayak.gayatri18@siesgst.ac.in

Tanvi Pongde, Student, Mumbai, India, pondge.tanvi18@siesgst.ac.in

Anindita Khade, Assistant Professor, Mumbai, India, anindita.khade@siesgst.ac.in

Abstract: Agreeing the 2010 worldwide weight of infection study, Persistent Kidney Illnesses (CKD) was positioned eighteenth in the rundown of reasons for complete number of passing around the world. 10% of the populace overall is influenced by CKD. The expectation of CKD can turn into an aid for the populace to foresee the wellbeing. Different strategy and procedures are going through the exploration stage for fostering the most exact CKD forecast framework. Utilizing Machine Learning procedures is the most encouraging one in this space because of its processing capacity and Machine Learning rules. This paper centers around a novel methodology for improving the forecast of CKD. In ongoing time Machine Learning has found its utilization in infection analyses, which is relied on expectation from side effects informational collection. Persistent kidney illness location framework utilizing Machine learning calculations is appeared here. This arrangement of CKD Discovery acknowledges upsides of traits as information and it is prepared and tried by different calculations.

Keywords: - ***Chronic Kidney Disease (CKD); Random Forest (RF); Gradient Boosting (GB); Decision Tree (DT); KNEARESRTNeighbors (KNN); Naïve Bayes(NB); Support Vector Machines (SVM); Machine Learning (ML); prediction*****.**

I. INTRODUCTION

Persistent kidney infection (CKD) is a huge general medical issue around the world, particularly for low-and medium-pay nations. Persistent kidney illness (CKD) implies that the kidney doesn't fill in true to form and can't effectively channel blood. About 10% of the populace overall experiences (CKD), and millions kick the bucket every year since they can't get moderate treatment, with the number expanding in the older. As per the Worldwide Weight Infection 2010, ongoing kidney sickness (CKD) has been raised as a significant reason for mortality worldwide with the quantity of passing expanding by 82.3% over the most recent twenty years. CKD, in its beginning phases, has no side effects; testing might be the best way to see whether the patient has kidney sickness. Early location of CKD in its underlying stages can assist the patient with getting powerful treatment and afterward forbid the movement to ESRD. It is contended that consistently, an individual that has one of the CKD hazards factors, for example, a family background of kidney disappointment, significant degree of haemoglobin, hypertension, diabetes, and numerous different factors needs to get checked. There are five phases of CKD, the most genuine one is stage 5 on the grounds that, at this stage, the kidneys can't do the vast majority of their capacities. It is hard to pinpoint the CKD phase of every persistent particularly at the beginning phases. It additionally causes a high chance of death inside a brief timeframe, a patient should be hospitalized and suitably restored. The most well-known reasons for kidney illness are diabetes and high hemoglobin level in body.

The sooner they think about having this sickness, the sooner they can get therapy. To bring issues to light and to energize the individuals who are generally vulnerable to the infection to play out the tests intermittently, we trust that the sickness can be recognized with the most un-potential tests and for minimal price. Along these lines, the target of this exploration is to give a compelling model to anticipate the CKD by least number of indicators.

The report surveys different exploration works that focus on the conclusion of CKD utilizing distinctive canny methods. Likewise, presents the dataset source and depiction lastly presents the system utilized for the expectation, including the information pre-processing steps and the demonstrating stage. The last segment of the report shows the aftereffects of the test and talks about the presentation of ML calculations in distinguishing CKD. At last, it incorporates the end and future work of this work.

II. LITERATURE SURVEY

To deal with the high-hazard state of CKD influencing a huge number of individuals all throughout the planet. This paper attempted to break down persistent kidney illness dataset utilizing two significant kinds of information examination, statistical and predictive analysis to make a 100% precise model dependent on ML calculation. The outcomes got can be a key for acquiring experiences from the dataset, determining the CKD status of the new patients and receiving great

 DOI : 10.35291/2454-9150.2021.0346

techniques for improving the security, proficiency, and nature of the consideration measures toward CKD infection. Machine Learning techniques are used in the field of medical science to provide an assistance to medical experts in performing more accurate diagnosis by making use of various combination of various computational sciences to find out a pattern in previous data used for training and then use these patterns in order to classify the test data in one of the possible categories. Various researchers have used Machine learning algorithms to classify and predict CKD status of a patient. They evaluated 12 classification techniques by applying them to CKD data, efficiency was calculated by comparing the prediction with the actual medical results. Various metrics are used for performance evaluation such as predictive accuracy, precision, sensitivity and specificity. The results indicate that decision-tree performed best with nearly the accuracy of 98.6%, sensitivity of 0.9720, precision and specificity of 1

In a study carried out by Sahil Sharma, Vinod Sharma and Atul Sharma the overall aim is to apply various algorithms to the available dataset such as decision tree, svm, artificial neural networks etc. which widely helps in handling data and provides the best predictive model. The authors evaluated 12 classification techniques by applying them to the chronic kidney disease dataset.

Marwa Almasoud, Tomas E Ward in their research achieved an accuracy of 99.1 according to F1- measure from Gradient Boosting classifier. Also, they found that hemoglobin has higher importance for both Random Forest and Gradient boosting in detecting CKD. Finally, their results are among the highest compared to previous studies but with less number of features reached so far.

"Survey for the Prediction of Chronic Kidney Disease using Machine Learning" by Pooja Sharma, Prof. Saket J Swarndeep.This study focuses on a novel approach for improving the prediction of CKD. In recent time Neural network system has discovered its use in disease diagnoses, which is depended upon prediction from symptoms data set. Chronic kidney disease detection system using neural network is shown here.

Bin Chen in his research proposed, a machine learning methodology for diagnosing CKD. After effectively filling out the incomplete data set, six machine learning algorithms (logistic regression, random forest, support vector machine, k-nearest neighbor, naive Bayes classifier and feed forward neural network) were used to establish models. Random Forest by using perceptron, which could achieve an average accuracy of 99.83% after ten times of simulation

Hanyu Zhang⁎, Che-Lun Hung†, William Cheng-Chung Chu‡ Ping-Fang Chiu§ presented a study where the main focus was on the classification models , that is, tree-based decision tree, random forest, and logistic regression has been analyzed. Preprocessing of the data is done, then filter method is used for feature selection that is univariate selection and correlation matrix along with feature importance to find best features from the dataset.

Later in 2020, Rahul Gupta,Nidhi Koli

, Niharika Mahor, N Tejashri proposed a study which elaborates the proposed system consisting of 4 main modules, which are data preprocessing, feature extraction, defining zones based on blood potassium level, diet recommendation module. The objective is to use machine learning algorithm and suggest suitable diet plan for CKD patient using classification algorithm on medical test records. Diet recommendation module is purely based on blood potassium level.

"Chronic Kidney Disease Prediction and

Recommendation of Suitable Diet plan by using Machine Learning" by Akash Maurya,Rahul Wable, Rasika Shinde, Sebin John presented a study in which using demographic data and medical care information of two population groups: people diagnosed with CKD in Colombia during 2018, and, a sample of people without a diagnosis of this disease.Using Artificial Neural Network (ANN) the model achieved 95% accuracy in the test dataset but as the machine-learning paradigm is opaque to the expert regarding the explanation of the outcome, we apply and validate a Neural Network – Case Based Reasoning (NN-CBR) twin system for the explanation of CKD predictions.

In 2020, Adeola Ogunleye and Qing-Guo Wang proposed "XGBoost Model for Chronic Kidney Disease Diagnosis". A study was done regarding medical diagnostic systems using Artificial Intelligence (AI) algorithms. Based on our study we concluded that applied methods of AI in healthcare provide beneficial results by improved diagnosis process and to detect the disease in early stages which follows to pick the suitable treatment plan. Major medical areas that were reviewed were related to cardiology, neurology, cancer, kidney disease, diabetics, cholera, and dental disease respectively using AI diagnostic criteria.

III. METHODOLOGY

i) Classifiers:

Decision Tree:

Decision tree classifiers classify data by making use of tree structure algorithms. The underlying algorithm begins with the training samples and corresponding class labels. The training set is partitioned recursively based on a feature value into subsets. Each internal node represents a test on attribute; each edge (branch) represents an outcome of the test. A decision tree classifier identifies the class label of an unknown sample by following path root to the leaves, which represent the class label for that sample. The feature (attribute) i.e., selected as the root node is the one that best divides the training data.

Support Vector Machine (SVM)*:*

These classifiers depend on primary danger minimization head and factual learning hypothesis with a point of deciding the hyperplanes (choice limits) that produce the effective

 DOI : 10.35291/2454-9150.2021.0346

division of classes. The basic calculation is Support Vector Classification (SVC) and it spins around the impression of a "margin"- on one or the other side of a hyperplane that partitions two information classes.

Augmenting the margin makes the biggest conceivable distance among the hyperplane and the occurrences on one or the other side of the hyperplane decrease an upper bound on the expected speculation blunder. It chips away at two kinds of information i.e., straightly divisible information and directly non-distinguishable information.

K-Nearest Neighbors*:*

The k-nearest neighbors classifier is amongst the simplest of all machine learning algorithms. It is based on the principal that the samples that are similar lie in close proximity. It classifies the test objects on the basis of number of closest training examples. It is also termed as a lazy- learning algorithm. KNN is a non-parametric algorithm which means that it does not assume anything on the underlying data distribution. In this, the Euclidean distance is calculated between the test data and every sample in the training data followed by classifying the test data into a class in which most of k-closest neighbors of training data belongs to. K is usually a very small positive integer.

As the value of K increases it becomes difficult to distinguish between the various classes. Cross-validation along with other heuristic techniques are used to choose an optimal value of K.

Naïve Bayes:

Naïve Bayes is a classification algorithm that is based on the Bayesian probability theorem. The classifier operates under the fundamental Naive Bayes assumptions which are independent and equal of feature contribution to the outcome, which means that feature presence or absence is unrelated to the presence or absence of any other feature. Naïve Bayes classifies an instance by calculating the probability of its belonging to each target class. In the rule of Naïve Bayes, A is the target class and B is the features vector describing an instance, P(A|B) is the probability of the instance B belonging to the target class, A. P(A) is the prior probability of the target class in the training set. P(B) is the probability of the features vector given the target class in the training set.

Random Forest:

Random Forest is a mainstream ML calculation that has a place with the administered learning procedure. It tends to be utilized for both Arrangement and Relapse issues in ML. It depends on the idea of ensemble learning, which is a cycle of joining different classifiers to take care of a complex problem issue and to improve the presentation of the model.

As the name proposes, "Random Forest is a classifier that contains various Decision trees on different subsets of the given dataset and takes the normal to improve the predictive precision of that dataset." Rather than depending on one decision tree, the Random Forest takes the forecast from each tree and dependent on the majority votes of expectations, and it predicts the last yield.

Gradient Boosting Classifier:

Gradient Boosting is an ML strategy for regression and classification problems, which creates an expectation model as an outfit of feeble forecast models, ordinarily decision trees. At the point when a decision tree is the weak learner, the subsequent calculation is called gradient boosted trees, which ordinarily outflanks random forest. It assembles the model in a stage wise design like other boosting techniques do, and it sums them up by permitting streamlining of a self-assertive differentiable loss function.

ii)System Design

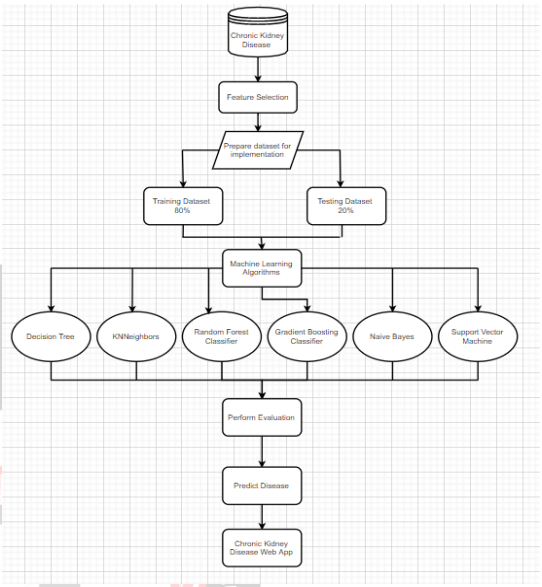

Fig. 1. Project Flow

The above figure shows the steps involved in executing the project.

iii) Chronic Kidney Disease Dataset:

The dataset that supports this research is based on CKD patients collected from Apollo Hospital, India in 2015 taken over a two-month period. The data is available in the University of California, Irvine (UCI) data repository named Chronic Kidney Disease Dataset . These data consisting of 400 observations suffer from missing and noisy value. The data includes 250 records of patients with CKD and 150 records of persons without CKD. Therefore, the percentage of each class is 62.5% with CKD an37.5% without CKD. The ages of these observations are varied from 2 to 90 years old. It can be seen from Table I that the CKD dataset has 24 features including 11 numeric features and 13 nominal features, and the 25th feature indicates the classification or state of CKD.

iv)Data Preprocessing

Data Reduction: Out of 25 attributes present in the dataset, we have selected 12 important attributes required to build predictive model. values in the dataset like NA's or blank values are removed by using function "ReplaceMissingValues", which

Replaces NAs with the mean values of that attribute.

 DOI : 10.35291/2454-9150.2021.0346

The dataset is divided into two sub datasets both containing 12 attributes.

Training dataset is derived from main dataset and it contains 320 out of 400 records in main dataset of CKD and testing dataset is 80 records

320 rows × 12 columns	80 rows × 12 columns

Table. 1. Description of Chronic Kidney Disease

Name	**Description**	Type:
Age (age)	Patient's age	Numeric
Blood Pressure (bp)	Blood pressure of the patient	Numeric
Specific Gravity(sg)	The ratio of the density of urine	Nominal
Albumin (al)	Albumin level in the blood	Nominal
Sugar (su)	Sugar level of the patient	Nominal
Red Blood Cells (rbc)	Patients' red blood cells count	Nominal
Pus Cell (pc)	pus cell count of patient	Nominal
Pus cell Clumps (pcc)	Presence of pus cell clumps in the blood	Nominal
Bacteria (ba)	Presence of bacteria in the blood	Nominal
Blood glucose(bgr)	blood glucose random count	Numeric
Blood Urea (bu)	blood urea level of the patient	Numeric
Serum Creatinine (sc)	serum creatinine level in the blood	Numeric
Sodium (sod)	sodium level in the blood	Numeric
Potassium (pot)	potassium level in the blood	Numeric
Hemoglobin (hemo)	hemoglobin level in the blood	Numeric
Packed cell volume (pcv)	packed cell volume in the blood	Numeric
White blood cell count (wc)	white blood cell count of the patient	Numeric
Red blood cell count(rc	red blood cell count of the patient	Numeric
Hypertension(htn)	Does the patient has hypertension on not	Nominal
Diabetes mellitus(dm)	Does the patient has diabetes or not	Nominal
Coronary artery disease	Does the patient has coronary artery disease or not	Nominal
Appetite(appet)	Patient's appetite	Nominal
Pedal Edema(pe)	Does patient has pedal edema or not	Nominal
Anemia(ane)	Does the patient have anemia or not	Nominal
Class	Does the patient has kidney disease or not	Nominal

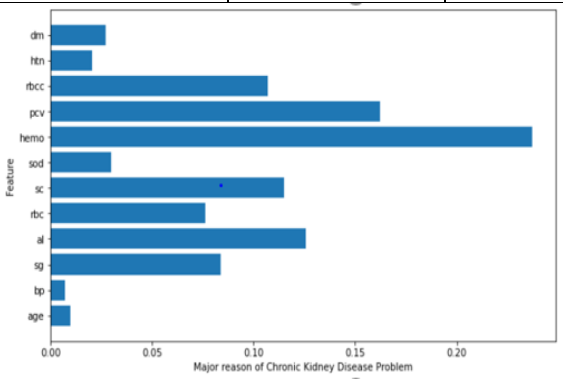

Fig. 2. Main Attributes that are reason of the disease.

IV. EXPERIMENTAL RESULTS

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience or instruction, in order to look for patterns in data and make better decisions in the future.

The Model of Chronic Kidney Disease Prediction System includes the execution of six different machine learning algorithms:

i) Decision Tree:

In the decision tree model, the confusion matrix for the dataset as follows. It states that the actual and predicted values for false negatives, there are 28 entries in the dataset that's are rightly predicted. Also, for true positive the rightly predicted entries are 51.

Accuracy: 98.80%

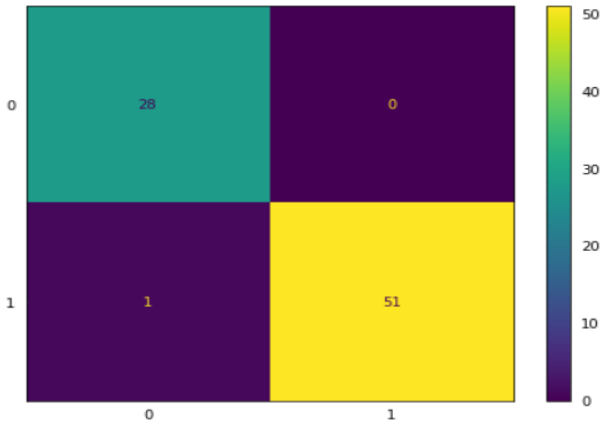

Fig. 3. Confusion Matrix for Decision Tree

ii) KNNeighbors:

In the KNNeighbors model, the data we use is standardized data and thus confusion matrix obtained is as follows. The number of false negatives is less than the true positives.

 DOI : 10.35291/2454-9150.2021.0346

Accuracy: 96.30%

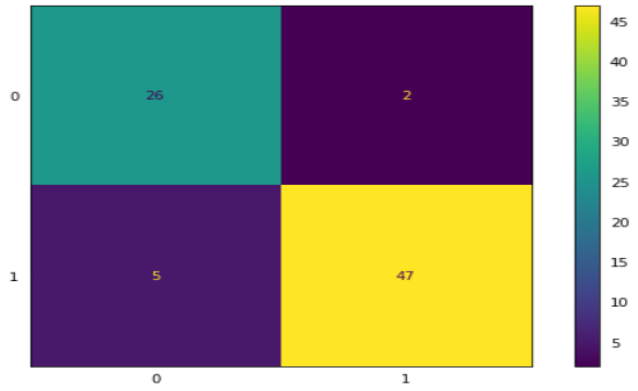

Fig. 4. Confusion Matric for KNNeighbors

Also, the best value of k (KNN coefficient) where we get the best score is given follows: -

```
k_range=range(1,30)
scores={}
h_score = 0
best_k=0
scores_list=[]

for k in k_range:
    knn=KNeighborsClassifier(n_neighbors=k)
    knn.fit(x_train,y_train)
    prediction_knn=knn.predict(x_test)
    scores[k]=accuracy_score(y_test,prediction_knn)
    if scores[k]>h_score:
        h_score = scores[k]
        best_k = k

    scores_list.append(accuracy_score(y_test,prediction_knn))
print('The best value of k is {} with score : {}'.format(best_k,h_score))
```

```
The best value of k is 5 with score : 0.9375
```

iii) Random Forest Classifier:

In the Random Forest Classifier model, the the confusion matrix for the dataset as follows. It states that the actual and predicted values for false negatives there are 28 entries in the dataset are rightly predicted. Also, for true positive the rightly predicted entries are 52

Accuracy: 99.9%

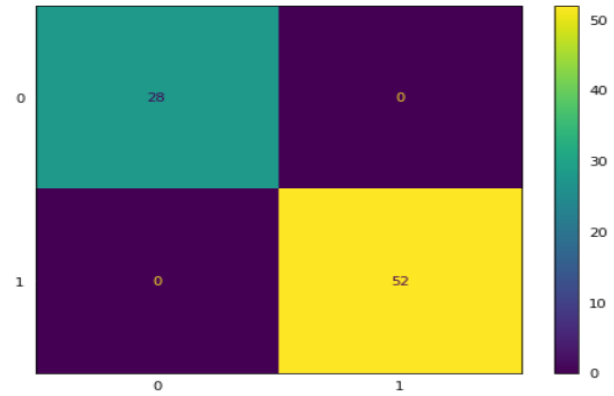

Fig. 5. Confusion Matrix for Random Forest

iii) Gradient Boosting Classifier:

In the Gradient Boosting Classifier model, the confusion matrix for the dataset as follows. It states that the actual and predicted values for false negatives is less than true positives. The entries rightly predicted for true positive are 52.

Accuracy: 99.7%

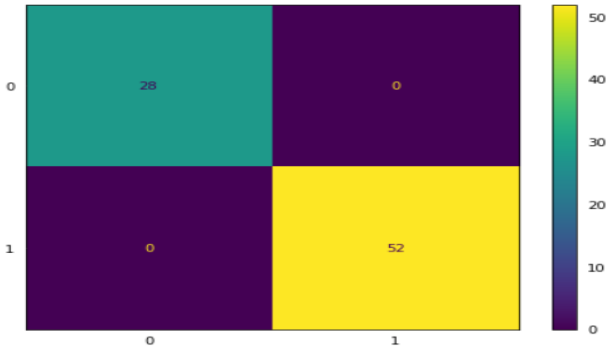

Fig. 6. Confusion Matric for Gradient Boosting

v) Naïve Bayes Classifier:

In the Naive Bayes Classifier model, the, the confusion matrix for the dataset as follows. It states that the actual and predicted values for false negatives,27 entries in the dataset are rightly predicted. Also, for true positive the rightly predicted entries are 49. There are 1 and 3 entries respectively which are not accurately predicted.

Accuracy: 95.00%

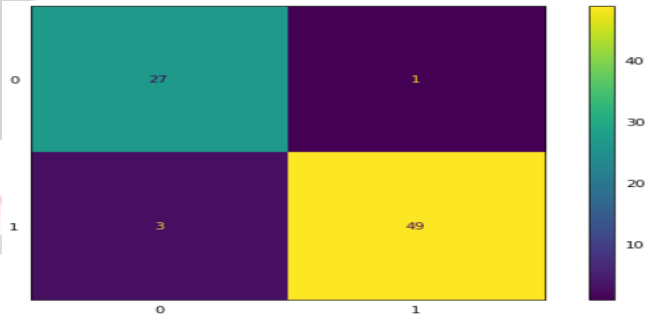

Fig. 7. Confusion Matric for Naïve Bayes

vi) Support Vector Machine Model:

In the Support Vector Machine model, the, the confusion matrix for the dataset as follows. It states that the actual and predicted values for false negatives is less than true positives. The entries rightly predicted for true positives are 51

Accuracy: 97.50%

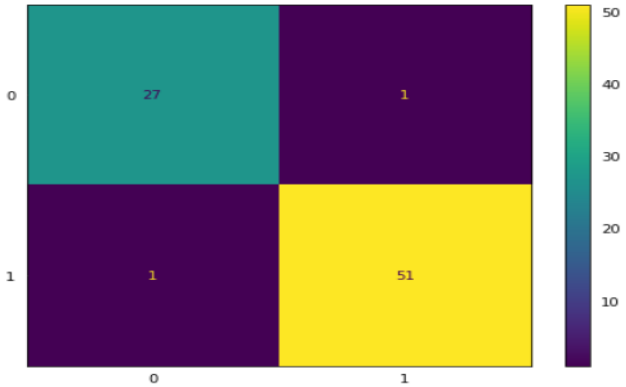

Fig. 8. Confusion Matric for SVM

We have got almost 100% accuracy in the testing data; hence we need to perform Hyperparameter Tuning to avoid overfitting.

 DOI : 10.35291/2454-9150.2021.0346

HYPERPARAMETER TUNING

```
from sklearn.ensemble import RandomForestClassifier
```

```
rfc_model = RandomForestClassifier()
```

```
# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 10, stop = 80, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [2,4]
# Minimum number of samples required to split a node
min_samples_split = [2, 5]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2]
# Method of selecting samples for training each tree
bootstrap = [True, False]
```

```
param_grid = {'n_estimators': n_estimators,
              'max_features': max_features,
              'max_depth': max_depth,
              'min_samples_split': min_samples_split,
              'min_samples_leaf': min_samples_leaf,
              'bootstrap': bootstrap}
print(param_grid)
```

```
{'n_estimators': [10, 17, 25, 33, 41, 48, 56, 64, 72, 80], 'max_features': ['auto', 'sqrt'], 'max_depth': [2, 4], 'min_samples_
split': [2, 5], 'min_samples_leaf': [1, 2], 'bootstrap': [True, False]}
```

```
from sklearn.model_selection import GridSearchCV
rf_Grid = GridSearchCV(estimator = rf_Model, param_grid = param_grid, cv = 3, verbose=2, n_jobs = 4)
```

```
x,y=kidney.loc[:,:'dm'],kidney['Class']
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=10,test_size=0.3,shuffle=True)
```

```
rf_Grid.fit(x_train, y_train)
```

```
Fitting 3 folds for each of 320 candidates, totalling 960 fits
```

```
[Parallel(n_jobs=4)]: Using backend LokyBackend with 4 concurrent workers.
[Parallel(n_jobs=4)]: Done  34 tasks      | elapsed:    1.9s
[Parallel(n_jobs=4)]: Done 504 tasks      | elapsed:   10.6s
[Parallel(n_jobs=4)]: Done 960 out of 960 | elapsed:   19.0s finished
```

```
GridSearchCV(cv=3, estimator=RandomForestClassifier(), n_jobs=4,
             param_grid={'bootstrap': [True, False], 'max_depth': [2, 4],
                         'max_features': ['auto', 'sqrt'],
                         'min_samples_leaf': [1, 2],
                         'min_samples_split': [2, 5],
                         'n_estimators': [10, 17, 25, 33, 41, 48, 56, 64, 72,
                                          80]},
             verbose=2)
```

```
rf_Grid.best_params_
```

```
{'bootstrap': False,
 'max_depth': 4,
 'max_features': 'auto',
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'n_estimators': 25}
```

```
print (f'Train Accuracy - : {rf_Grid.score(x_train,y_train):.3f}')
print (f'Test Accuracy - : {rf_Grid.score(x_test,y_test):.3f}')
```

```
Train Accuracy - : 0.993
Test Accuracy - : 1.000
```

Fig. 9. Hyperparamter Tuning for Random Forest

After Hyperparameter Tuning the Test Accuracy is 100%. We can conclude that random forest classifier is best predictive algorithm for the chronic kidney disease dataset.

vii) Auroc Values & Roc Graph:

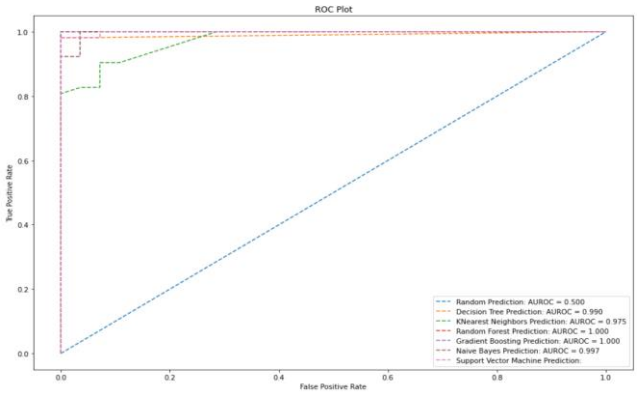

Fig. 10. ROC Graph

vii) DEPLOYING THE WEB PORTAL FOR CHRONIC KIDNEY DISEASE PREDICTION:

A web portal has been developed for the easy use of the chronic kidney disease portal which will feasible for both medical professionals and also common people. The portal consists of all the input values that is the 12 attributes that we have considered for our project and it is been developed using the most effective algorithm that is Random Forest Algorithm. The portal predicts whether a person has chances of having chronic kidney disease or not.

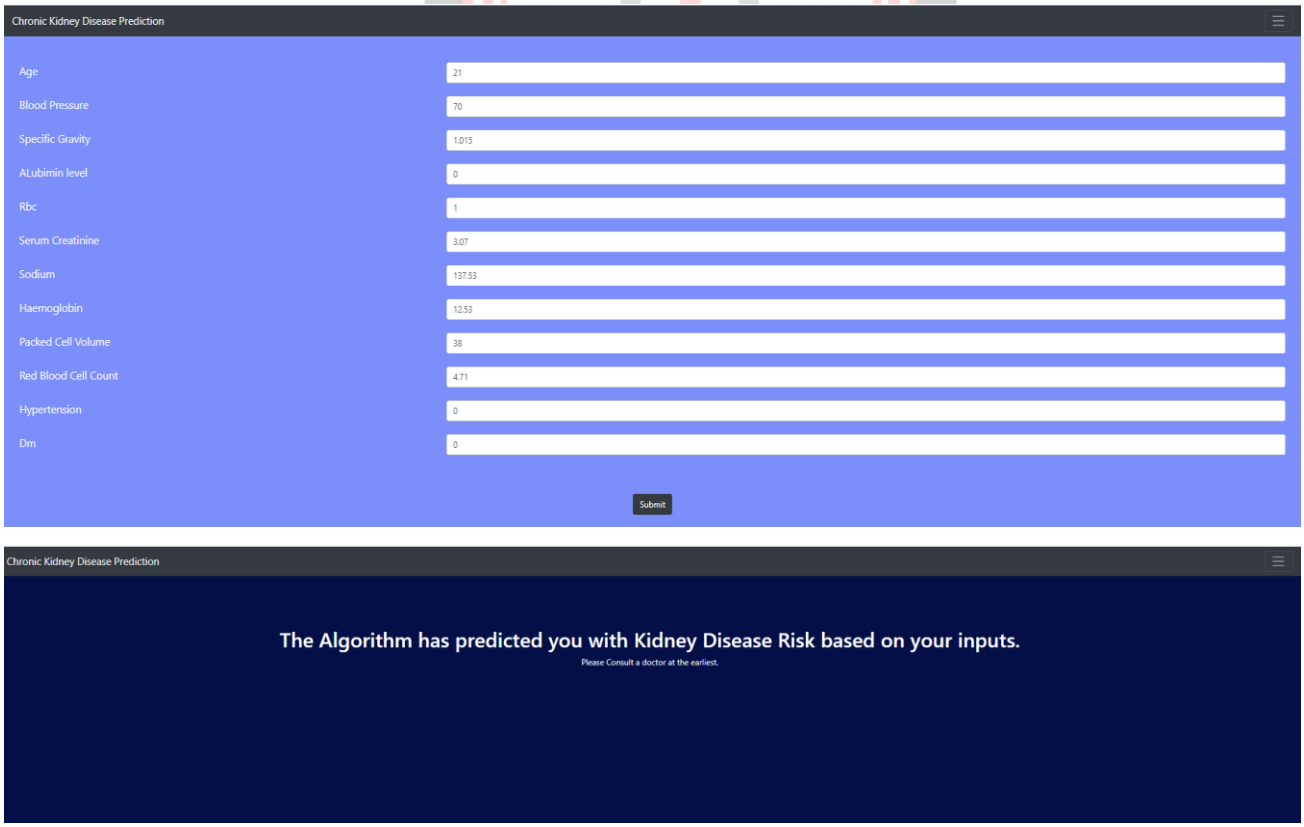

Fig. 11. Output for 'CKD' = True

In this GUI, a HTML form is presented to the user wherein the user can enter the required attributes for predicting whether the user is at risk of suffering from a chronic kidney disease or not. In the above example, the output can be seen to be '**The**

 DOI : 10.35291/2454-9150.2021.0346

algorithm has predicted the user with Ckd based on the inputs', that is the user is identified to be at a risk of suffering from a chronic kidney disease.

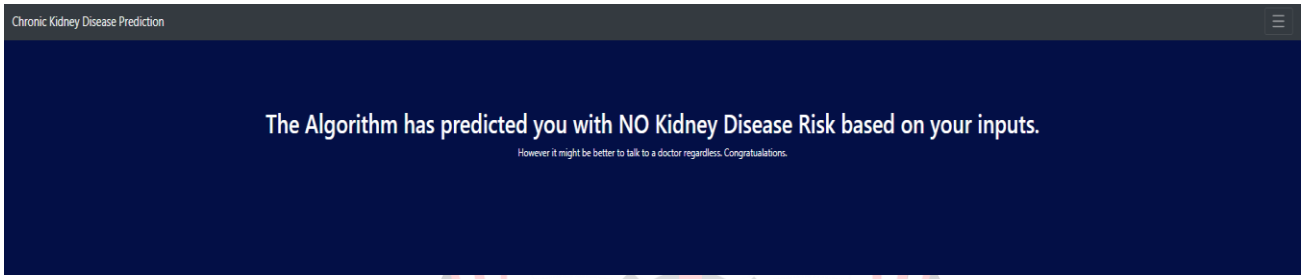

Figure No.12 - Output for 'CKD' = False

In the above example, the output shows that the user is not at a risk of suffering from a chronic kidney disease and the output is '**The algorithm has predicted the user with Ckd based on the inputs**' that is the user is identified to be at a risk of suffering from a chronic kidney disease.

V. CONCLUSION AND FUTURE SCOPE

In this project we have studied different machine learning algorithms. We have analysed 12 different attributes related to CKD patients and predicted accuracy for different machine learning algorithms like Decision tree, KNN, Random Forest, Gradient Boosting, Naïve Bayes, and Support Vector Machine. From the results analysis, it is observed

CLASSIFIERS	ACCURACIES	AUC
DECISION TREE	98.80%	99.00%
KNNEIGHBORS	96.30%	97.50%
RANDOM FOREST	99.9%	100%
GRADIENT BOOSTING	99.7%	100%
NAÏVE BAYES	95.00%	99.70%
SVM	97.50%	99.90%

This work examines the ability to detect CKD using machine learning algorithms while considering the least number of tests or features. We approach this aim by applying four machine learning classifiers: logistic regression, SVM, random forest, and gradient boosting on a small dataset of 400 records. In order to reduce the number of features and remove redundancy, the association between variables have been studied. A filter feature selection method has been applied to the attributes and found that there are hemoglobin, albumin, and specific gravity have the most impact to predict the CKD.

Limitations of this study are the strength of the data is not higher because of the size of the data set and the missing attribute values. To build a machine learning model targeting chronic kidney disease with overall accuracy of 100%, will need millions of records with zero missing values. This work will be considered as basement for the healthcare system for CKD patients. Also, extension to this work is that implementation of deep learning since deep learning provides high-quality performance than machine learning algorithm.

The proposed system will definitely help in improving the prediction of chronic kidney disease system by increasing its accuracy and prediction capability by reducing the error. The advantage of this system is that, the prediction process is less time consuming. It will help the doctors to start the treatments early for the CKD patients and also it will help to diagnose more patients within a less time period.

 DOI : 10.35291/2454-9150.2021.0346

REFERENCES

[1] "Diagnosis of Chronic Kidney disease using effective classification and feature selection technique "– Nusrat Tazin, Shahed Anzrus Sahab, Muhammed Tawfiq Choudhary, 2016.

[2] "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD)" – K.D.M Perera, 2017.

[3] "Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis" – Sahil Sharma, Vinod Sharma, Atul Sharma ,2017.

[4] "Pre-emptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques" - Reem A. Alassaf, Khawla A. Alsulaim, Noura Y. Alroomi - 2018

[5] "Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors" - Marwa Almasoud, Tomas E Ward – 2019

[6] "Survey for the Prediction of Chronic Kidney Disease using Machine Learning" - Pooja Sharma, Prof. Saket J Swarndeep, 2019.

[7] "Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning" - Gabriel R. Vásquez-Morales1, Sergio M. Martínez-Monterrubio2, Pablo Moreno-Ger3, And Juan A. Recio-García,2019

[8] "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease" - Jiongming Qin 1, Lin Chen, Yuhua

[9] "Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients" - Erik Dogvan, Mohy Uddin, Yu-chan-Li, Shabbir Syed Abdul ,2020.

[10] "XGBoost Model for Chronic Kidney Disease Diagnosis" - Adeola Ogunleye and Qing-Guo Wang,2020

[11] "Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease" - Rahul Gupta, Nidhi Koli, Niharika Maho, N Tejashri,2020

[12] "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using Machine Learning" - Akash Maurya,Rahul Wable,Rasika Shinde ,Sebin John ,Rahul Jadhav,Dakshayani.

[13] "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks" - Hanyu Zhang, Che-Lun Hung, William Cheng-Chung Chu Ping-Fang Chiu and Chuan Yi Tang

 DOI : 10.35291/2454-9150.2021.0346