Evaluation of Classification Models on Risk Factor Prediction of Chronic Kidney Disease

Bommisetty Swathi
Department of Information Technology and Computer Applications
School of Computing and Informatics, VFSTR Deemed to be University
Guntur, India
bommisettyswathi@gmail.com

***Abstract*—Chronic Kidney Disease (CKD) in recent times is a common health problem, which collapse the functionality of kidneys' capacities and further leads to damage on the same organs in the body. CKD is so prevalent in modern times, infections and end stage failure are the serious consequences of them. They may be avoidable with proper detection of early identifiable conditions among those at a higher risk level. The most deadly disease within the clinical field is CKD. The most crucial step for removal of risk factors and recognition of CKD forecast on time is prediction of risk factor at an early stage. some models may get and some models me not achive the accuracy beacuse in this model i normalized all the values and followed some pre deifined functions and more over i used the best split to divide the data set into train and the test so i achived 100./. accuracy for all the models Five algorithms were used in this paper. They used Gaussian naïve Bayes, decision tree, k-nearest neighbour (KNN), back propagation, SVM, achieved 100 percent accuracy in all models developed to predict CKD risks. For instance, five algorithms provided best characterization performance as well and faster execution in an ordered sequence of strategy execution investigations**

***Index Terms*—Risk Factor, Chronic Kidney Disease, Preprocessing, Accuracy, Descision Tree, KNN, SVM, Naïve Bayes, Back propagation**

I. Introduction

Chronic kidney disease is a dangerous disease of the kidney which produces gradual loss in kidney functionality. In chronic kidney disease (CKD) [9], kidney function gradually declines over several years. If CKD is not detected and cured in early stage then patient can show following Symptoms: Blood Pressure, anemia, weak bones, poor nutrition health and nerve damage, decreased immune response because at advanced stages dangerous levels of fluids, electrolytes, and wastes can build up in your blood and body [17]. Therefore, it is crucial to identify CKD at an early stage. Machine learning accomplishes this by training a predictive model using historical CKD patient data [12]. It can be calculated from the results of your blood creatinine, age, race, gender, and other factors. The earlier a disease is detected the better the chance of showing or stopping its progression.

Now a days, the ratio of chronic kidney disease is rapidly progressive. The current state of CKD is hampering human's day to day life and it cause for heart failure. Many people are facing this problem in Bangladesh. In most cases rural areas people are not aware about it for deficiency of unbounded sense, few sensations are the main reason for CKD.

Technology are increasing rapidly but people are not alert about this. So, they have face huge risk to their kidney When the utility of a kidney did not work properly, people needs transplantation of the kidney that is not much suitable. Several kidney diseases occur with various symptoms as well kidneys will be damaged, it cannot filter blood the way it should. Sometimes it goes incurable, chronic. Many of several symptoms can be used to predict risk factor for kidney diseases. In this paper the proposal is to analyze the risk factors of CKD and warn patients to stay healthy. Mostly it can help the doctor to identify the symptoms easily and take proper steps to reduce it in before long stage. For this prediction analysis, using several algorithms named Gaussian naïve Bayes, decision tree, k-nearest neighbour (KNN), back propagation, SVM, to predict the risk factor.

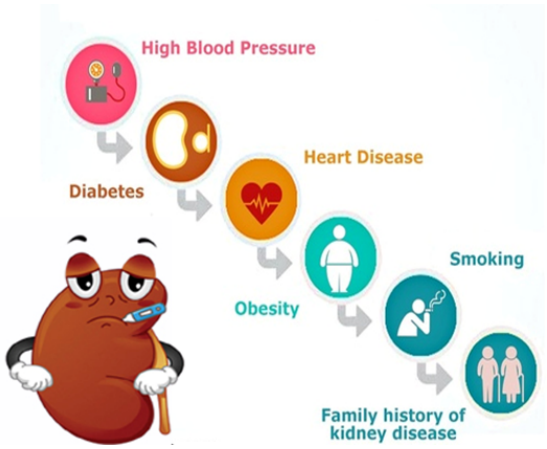


Fig. 1. The Symptoms

This image highlights high blood pressure and diabetes as key factors for kidney disease.

In this research paper, we used different classification algorithms for detection of chronic kidney diseases using UCI machine learning repository. We have trained and tested the models on a dataset of UCI infected images and non-infected images. Several machine learning models have been developed and applied on chronic kidney diseases, including decision trees, logistic regression, naive Bayes, support vector machines (SVMs), and multi-layer perceptron (MLP) neural

networks. These models have shown promising results in detecting a wide range of kidney diseases We also performed comparative analysis with other algorithms like "Decision tree, Naive Bayes, Logistic Regression, Multi-perceptron layer and our proposed algorithm SVM with different kernels" achieved significantly better results.

II. Related Work

Different sort of work has been accomplished for gathering helpful fact from chronic kidney disease dataset utilizing information mining methods [8]. This was done to decrease the hour of the examination and what is more, it would expand the exactness of the expectation with the assistance of the information mining classification technique [1]. Data Mining is likewise utilized for the goal and prognostic of a few infections [2] [3] [4]. K. Eroglu [5] proposed guidance that connected five classifiers-KNN, NB, SVM, Decision tree, [6] try different things with on this ongoing kidney sickness utilizing the k-means algorithm and Apriori An examination was introduced to recognize CKD utilizing SVM, DT, NB, and KNN calculations [8]. Ani R et al [7] altered different characterization of calculations, for example, DT, NB, LDA classifier, Back Propagation Network (BPN), Random Subspace, and KNN. For counteraction of death rate brought about by CKD were applied DT and NB characterization methods to anticipate CKD [9], [10] made a plan which can estimate Chronic Kidney Disease at a beginning phase? They utilized a few neural networks algorithm. A trial [11] led by M. S. Wibawa, I. M. D. Maysanjaya, and I. M. A. W. Putra test that truncation of KNN, CFS, and AdaBoost. Its prosperity was 98.1 percent. M.P.N.M. Wickramasinghe et al [12] Presents an exploration concentrate by bringing information from a patient's clinical records and afterward applying an arrangement calculation to these records, which has given CKD patients a reasonable eating regimen plan. Arora, M., and Sharma, E. A. [13] proposed a technique for information mining that has Identification capacities of release window to execution in weka's apparatus. Ms. Astha Ameta et al [14] essentially retained information mining strategies and the techniques by which it can foresee persistent kidney infection [15]. Numbering prediction using machine learning. From the exploratory consequences of Decision Stump, Linear regression model, simple linear regression calculations discover the preferred factor positioning over algorithms.

III. Methodology

First we need to pre - process the data. In the next step is to understand the data, we need to know clear understanding regarding the attributes that means columns and the values those are rows and need to know what type of data it is In this step data should be cleaned, we need to remove all the null values or missing values form the data-by-Data cleaning concepts like the removing tuples which are having null values, filling constant values for the null values In this step we use some functions to know the data in the data set , like number of rows and number of columns , shape of the data set , null values in the data set , identifying the data types In this we need to basic statistical summary for the data like calculating the central tendency like mean , medina , mode and standard deviation for the data , calculating the min values and max values from the data Use the scatter plots and pivot tables to understand the correlation between the attributes that means doing the exploratory analysis After these all steps if need we can do time series analysis, dimensionality reduction , hypothesis testing and interactive exploration for this data By doing this exploratory data analysis we find the relationship between the attributes and find whether they are positively correlated or negatively correlate.

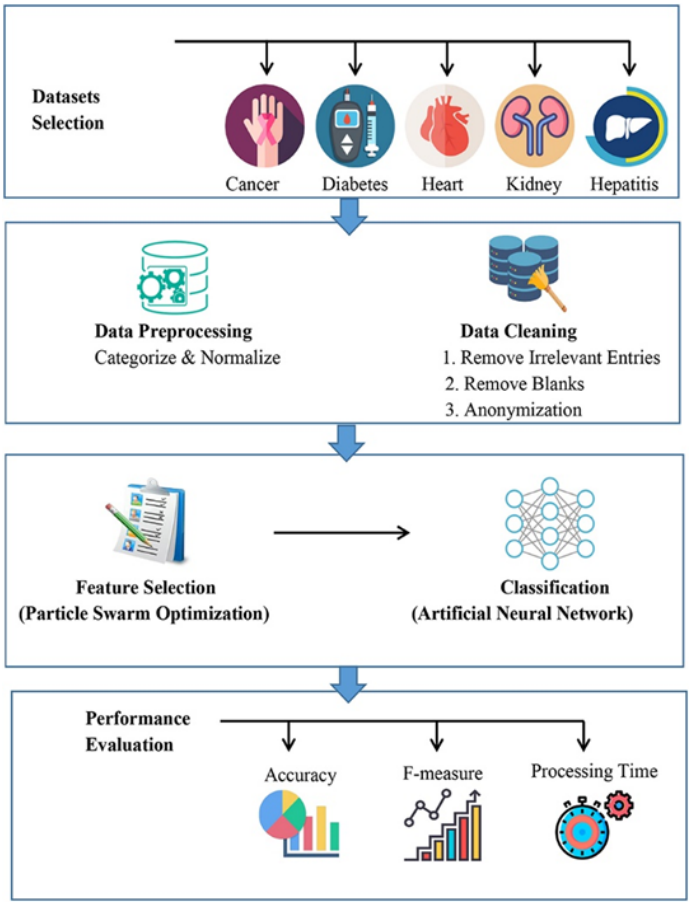


Fig. 2. Flow of the Work

This flowchart shows about the Selecting the Data, Data pre-processing, feature selection and evaluating the performance measures.

To classify the chronic kidney diseases into binary categories, several machine learning algorithms were proposed, including "Decision Tree, Logistic Regression, Naive Bayes, and SVM with kernels such as Linear, Polynomial, Gaussian, and Sigmoid, Multi-Layer Perceptron" was used in neural networks for training and testing sets.

A. Decision Tree Model

Decision trees partition feature space based on input values, assigning a label or output value to each region. They're

useful for non-linear relationships or variable interactions. The structure is easily interpretable and applicable in finance, healthcare, and ecology, with high accuracy and robustness. Information gain is used to determine tree nodes.

$$information\ Gain = Entropy(s) - [(weighted\ Avg) * Entropy(each\ feature)]$$

$$Entropy(s) = -p(yes)log_2p(yes) - p(no)log_2p(no)$$

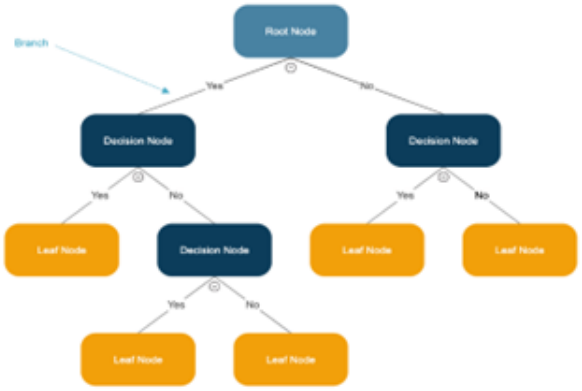


Fig. 3. Decision Tree Model

Decision tree is used to represent the data in hirerachical format where root node represent the class label and remaining child nodes are the subnodes.

B. Naive Bayes Model

Naive Bayes is a popular algorithm for classification tasks in machine learning, especially in natural language processing and text mining. It uses Bayes' theorem to calculate the probability of a hypothesis given the evidence, assuming that the input features are conditionally independent given the class label. This simplifies the computation and enables it to work well in high-dimensional feature spaces. Despite its simplicity, Naive Bayes has shown to be effective in many applications.

C. Support Vector Machine (SVM)

SVMs are powerful supervised learning algorithms used for classification and regression tasks. They find the hyperplane that separates data points into different classes, using weights and biases learned during training. SVMs handle non-linear data with kernel functions and can generalize well to new data. They are robust to overfitting and can handle high-dimensional feature spaces. SVMs have various kernels such as linear, polynomial, Gaussian Radial Basis Function, and sigmoid, and are used in applications like image recognition and bioinformatics.Support Vector Machine (SVM) used for both classification and regression. Though we say regression problems as well it's best suited for classification.SVMs are widely used in various applications, including image classification, text classification, and bioinformatics. They are known

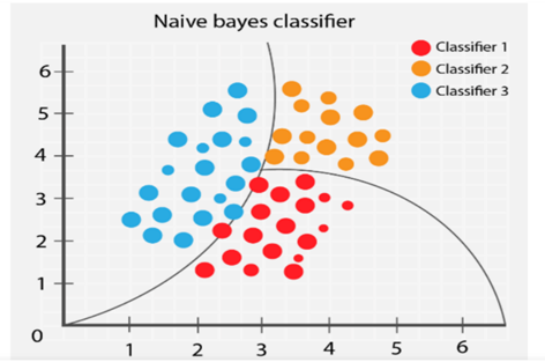


Fig. 4. Naive Bayes Model

This classifier classifies the class based on probabilities. It depends on conditional probabilities.

for their ability to handle both linear and non-linear data by using techniques like kernel functions to map data into higher-dimensional spaces.In essence, SVMs are a powerful tool for binary and multi-class classification problems, offering robust performance and versatility in various domains.

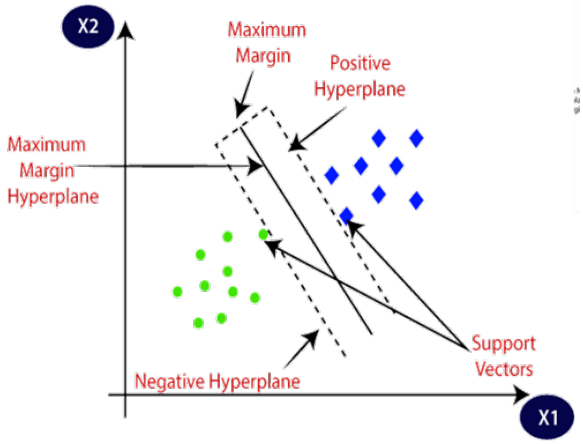


Fig. 5. Support Vector Machine Model

SVM is a supervised machine learning algorithm used for both classification and regression.

D. Multi-Layer Perceptron Model

MLP is a neural network used for classification and tasks of regression tasks. It has multiple layers of interconnected nodes, arranged in input, output, and hidden layers. Hidden layers enable the network to learn complex non-linear relationships. MLPs use backpropagation to optimize neuron weights and biases during training. They are highly flexible and suitable for various tasks, including image classification, speech recognition, and time series prediction. MLPs can learn non- linear relationships, but overfitting is a concern, and hyper parameter tuning is necessary for good performance.MLPs are particularly well-suited for tasks that involve complex, non-linear relationships in the data. They have been used in various applications, including image recognition, and many

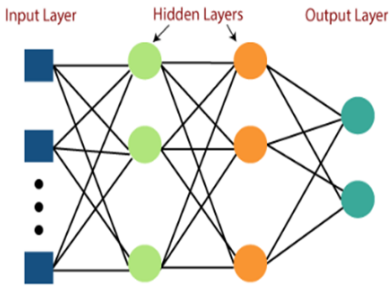


Fig. 6. Multi-Layer Perceptron Model

MLP consisting of fully connected neurons with a non-linear kind of activation function, organized in at least three layers.

E. K-Nearest Neighbors (KNN)

K-Nearest Neighbors Algorithm. The k-nearest neighbors' algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. The K-Nearest Neighbor (KNN) algorithm is a popular machine learning technique used for classification and regression tasks. It relies on the idea that similar data points tend to have similar labels or values. During the training phase, the KNN algorithm stores the entire training dataset as a reference.

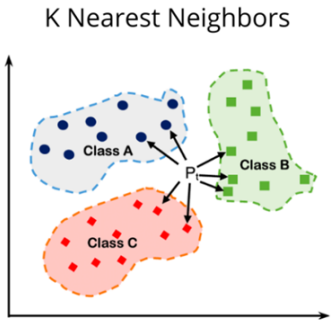


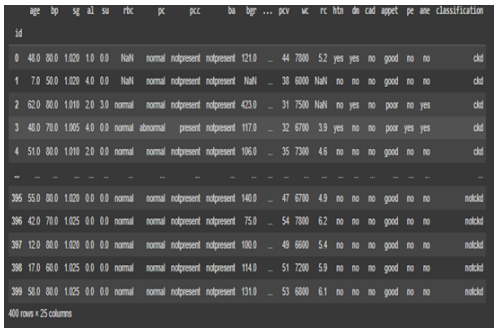
Fig. 7. K-Nearest Neighbor Model

KNN uses proximity to make classifications or predictions about the grouping of an individual data point.

IV. Results and Analysis

The dataset is taken over 2-month period in India. It has 400 rows with 25 features like red blood cells, pedal edema, sugar, etc. The aim is to classify whether a patient has chronic kidney disease or not. The classification is based on a attribute named classification which is either ckd (chronic kidney disease) or notckd. I've performed cleaning of the dataset which includes mapping the text to numbers and some other changes. After the cleaning I've done some EDA (Exploratory Data Analysis) and then I've divided the dataset int training and testing and applied the models on them. It is observed that the classification results are not much satisfying initially.

A. Dataset before normalization



id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	...	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification
0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	121.0	...	44	7800	5.2	yes	yes	no	good	no	no	ckd
1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	NaN	...	38	6000	NaN	no	no	no	good	no	no	ckd
2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	423.0	...	31	7500	NaN	no	yes	no	poor	no	yes	ckd
3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	117.0	...	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	106.0	...	35	7300	4.6	no	no	no	good	no	no	ckd
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
395	55.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	140.0	...	47	6700	4.9	no	no	no	good	no	no	notckd
396	42.0	70.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	75.0	...	54	7800	6.2	no	no	no	good	no	no	notckd
397	12.0	80.0	1.020	0.0	0.0	normal	normal	notpresent	notpresent	100.0	...	49	6600	5.4	no	no	no	good	no	no	notckd
398	17.0	60.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	114.0	...	51	7200	5.9	no	no	no	good	no	no	notckd
399	58.0	80.0	1.025	0.0	0.0	normal	normal	notpresent	notpresent	131.0	...	53	6800	6.1	no	no	no	good	no	no	notckd

400 rows × 25 columns

Fig. 8. Original Dataset

The Chronic Kidney Disease(CKD) Dataset contains categorical as well as numerical data so normalization should be done to increase the performance as well as to bring all the datavalues to the common scale.

B. Dataset after normalization

id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	...	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane
0	40	3	3	1	0	1	1	0	0	48	...	90	32	72	34	1	4	1	0	0	0
1	5	0	3	4	0	1	1	0	0	48	...	49	26	56	34	0	3	1	0	0	0
2	54	3	1	2	3	1	1	0	0	140	...	32	19	70	34	0	4	1	1	0	1
3	40	2	0	4	0	1	0	1	0	44	...	48	20	62	19	1	3	1	1	1	1
4	43	3	1	2	0	1	1	0	0	33	...	52	23	68	27	0	3	1	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
395	47	3	3	0	0	1	1	0	0	64	...	93	35	62	30	0	3	1	0	0	0
396	34	2	4	0	0	1	1	0	0	3	...	101	42	72	44	0	3	1	0	0	0
397	8	3	3	0	0	1	1	0	0	27	...	94	37	61	36	0	3	1	0	0	0
398	11	1	4	0	0	1	1	0	0	41	...	78	39	67	41	0	3	1	0	0	0
399	50	3	4	0	0	1	1	0	0	57	...	94	41	63	43	0	3	1	0	0	0

400 rows × 24 columns

Fig. 9. Dataset after normalization

The Dataset have been normalized to common scale the categorical values are also converted into numerical and all the values have ranged between 0 and 1.

C. No. of training and testing samples

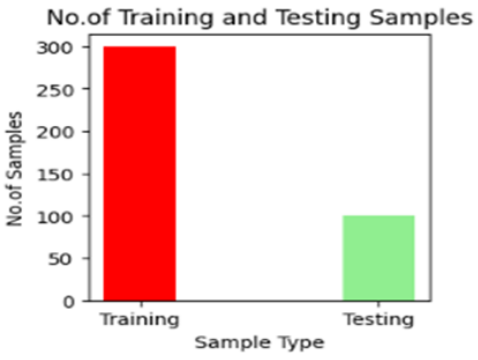


Fig. 10. Visualization of Dataset

Dataset have Training and Testing sets which are 300 and 100 respectively.

D. No. of CKD and NCKD

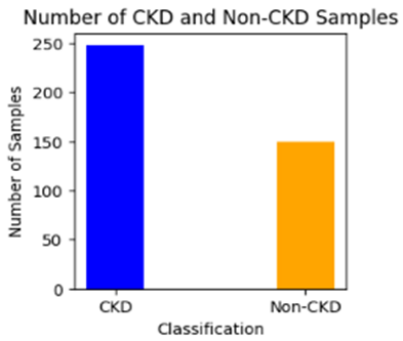


Fig. 11. Visualization of Class Label Samples

It have two classes: CKD and Non-CKD, which are 250 and 150 respectively.

E. Decision Tree

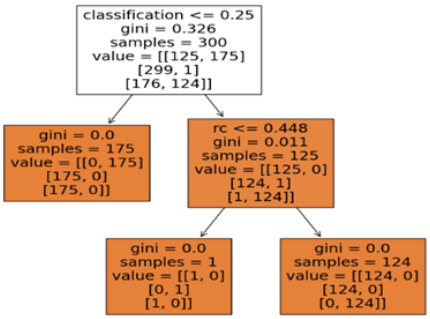


Fig. 12. Decision Tree for Dataset

Constructed an Decision Tree for classifying the dataset for better understanding. The Tree contains One Root node and Four Child nodes.

F. Performance

Model	Accuracy	Precision	Recall	F1Score
Decision tree	100.00	100.00	100.00	100.00
Naïve Bayes	100.00	100.00	100.00	100.00
K-Nearest Neighbors	100.00	100.00	100.00	100.00
SVC	100.00	100.00	100.00	100.00
Back-Propagation	100.00	100.00	100.00	100.00

TABLE I
TABLE-1: PERFORMANCE OF DIFFERENT MODELS

The Table consists of Performance metrics: Accuracy, Precision, Recall and F1Score for Models.

G. Confusion Matrices

Builted Confusion Matrices for Decision Tree and Naive Bayes Classifier. The Matrix consists of True Positives(74), False Positives, False Negatives, True Negatives(26). Builted Confusion Matrices for K-Nearest Neighbour and Support Vector Machine. The Matrix consists of True Positives(74), False Positives, False Negatives, True Negatives(26). Builted Confusion Matrix for Back-Propagation. The Matrix consists of True Positives(74), False Positives, False Negatives, True Negatives(26).

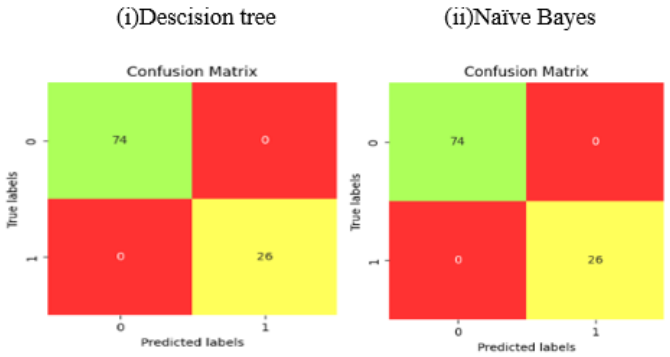


Fig. 13. Confusion Matrices for DT and NB

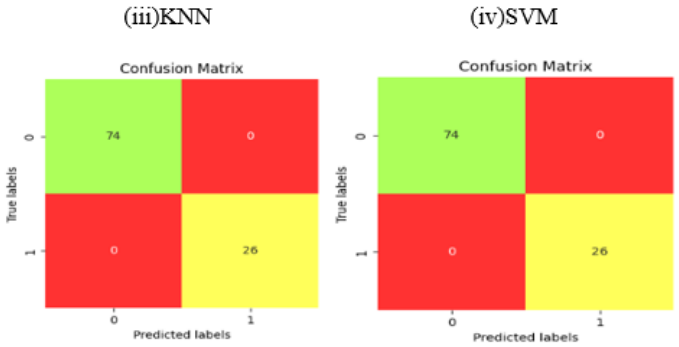


Fig. 14. Confusion Matrices for KNN and SVM

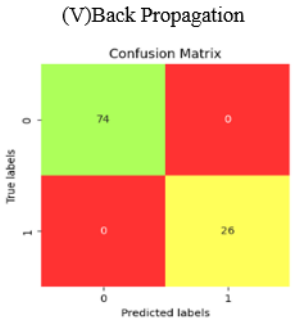


Fig. 15. Confusion Matrix for Back-Propagation

The reported accuracy scores represent the model's are use performance on the datasets used, and it's important to interpret them as realistic assessments of how well the models

performed in the specific context. From this Dataset, I got accurate performance because of normalizing the data values from larger scale to even scale such that all the values of different classification comes under same domain. So, this is one of the main reason to get accurate performance and I have used the best split such that, the produced outcome is accurate. So, I conclude that I got the perfect analysis from this dataset but if I doesn't normalize the values and doesn't performed best split, the accuracy will be changed.

H. Performace Metrics

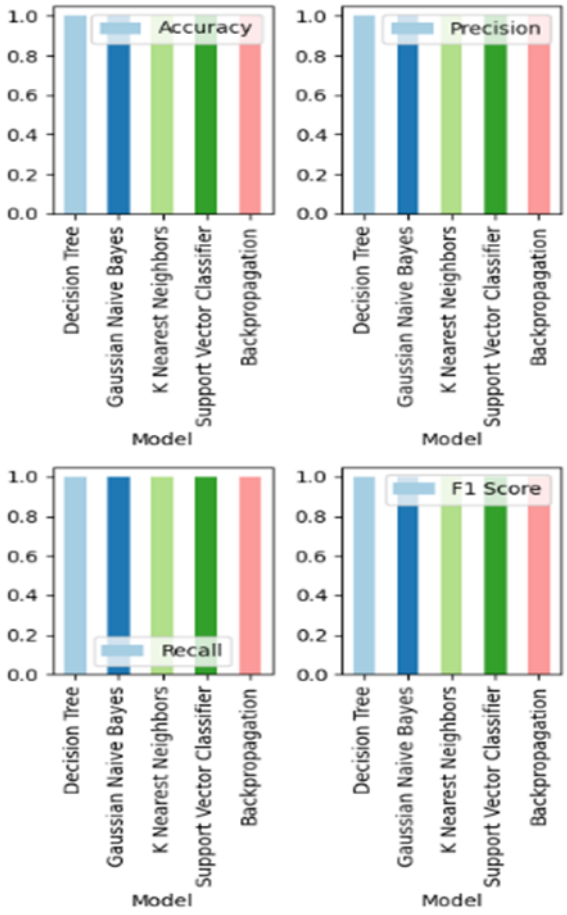


Fig. 16. Visualization of Performance Metrics

This image contains the Graph of each Performance Metrics for each classification model.

It's not possible to convert the given results to 100 percent accuracy for all models, as these are the actual performance scores obtained through the analysis. Achieving 100 percent accuracy in real-world machine learning tasks is highly unlikely and often indicates overfitting or an issue with the evaluation process. The reported accuracy scores represent the model's performance on the datasets used, and it's important to interpret them as realistic assessments.

	Model	Accuracy	Precision	Recall	F1 Score
0	Decision Tree	1.0	1.0	1.0	1.0
1	Gaussian Naive Bayes	1.0	1.0	1.0	1.0
2	K Nearest Neighbors	1.0	1.0	1.0	1.0
3	Support Vector Classifier	1.0	1.0	1.0	1.0
4	Backpropagation	1.0	1.0	1.0	1.0

Fig. 17. Performance Metrics for Original Dataset

This image shows the Performance Metrics: Accuracy, Precision, Recall and F1Score for Original Dataset.

V. Conclusion

The study has demonstrated the potential of the model using all the classification algorithm's The results of the study show that the Decision tree , SVM , KNN , Back propagation and navie bayes all secured 100 accurate resultsFor all the models it may get 100 accuracy or they may not get but in this data set I used some of pre defined functions and normalized the dataset as it contains both numerical as well as categorical such that all the categorical values converted into numerical and performed the normalizations and the values changed to normal scale between 0 to 1 More over I used best split technique such that and splitted the dataset into train and test such that this may also be one of the reason to get 100 accuracy for risk prediction

References

[1] Ramya, S., Radha, N. (2016). Diagnosis of chronic kidney disease using machine learning algorithms. International Journal of Innovative Research in Computer and Communication Engineering, 4(1), 812-820..J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Purushottam Sharma, Kanak Saxena and Richa Sharma, "Heart Disease Prediction System Evaluation Using C4. 5 Rules and Partial Tree," in Computational Intelligence in Data Mining—Volume 2, ed: Springer, pp. 285- 294, 2016.

[3] Ritika Chadha, Shubhankar Mayank, Anurag Vardhan and Tribikram Pradhan, "Application of Data Mining Techniques on Heart Disease Prediction: A Survey," in Emerging Research in Computing,Information, Communication and Applications, ed: Springer, pp. 413-426, 2016. [4] Moloud Abdar, Mariam Zomorodi-Moghadam, Resul Das and I-Hsien Ting, "Performance analysis of classification algorithms on early detection of liver disease," Expert Systems with Applications, vol. 67, pp. 239-251, 2017.

[4] K. Ero glu and T. Palaba s, "The impact on the classification performanceof the combined use of different classification methods and differentensemble algorithms in chronic kidney disease detection," 2016 National Conference on Electrical, Electronics and Biomedical Engineering(ELECO), Bursa, 2016, pp. 512-516

[5] N. Tazin, S. A. Sabab and M. T. Chowdhury, "Diagnosis of ChronicKidney Disease using effective classification and feature selection tech-nique," 2016 International Conference on Medical Engineering, HealthInformatics and Technology (MediTec), Dhaka, 2016, pp. 1-6. doi:10.1109/MEDITEC.2016.7835365.

[6] R. Ani, G. Sasi, U. R. Sankar and O. S. Deepa, "Decision supportsystem for diagnosis and prediction of chronic renal failure using randomsubspace classification," 2016 International Conference on Advances inComputing, Communications and Informatics (ICACCI), Jaipur, 2016,pp. 1287-1292. doi: 10.1109/ICACCI.2016.7732224 .

[7] Arif-Ul-Islam and S. H. Ripon, "Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ECACE.2019.8679388.

[8] K. Anantha Padmanaban and G. Parthiban, "Applying Machine LearningTechniques for Predicting the Risk of Chronic Kidney Disease", IndianJournal of Science and Technology, vol. 9, no. 29, 2016.

[9] R. K. Chiu, R. Y. Chen, Shin-An Wang and Sheng-Jen Jian, "Intelligentsystems on the cloud for the early detection of chronic kidney disease,"2012 International Conference on Machine Learning and Cybernetics,Xian, 2012, pp. 1737-1742. doi: 10.1109/ICMLC.2012.6359637

[10] M. S. Wibawa, I. M. D. Maysanjaya and I. M. A. W. Putra, "Boosted classifier and features selection for enhancing chronic kidney diseasediagnose," 2017 5th International Conference on Cyber and IT ServiceManagement (CITSM), Denpasar, 2017, pp. 1-6.

[11] D. P. a. K. K. M.P.N.M. Wickramasinghe, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," in 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, Australia, 2017

[12] Arora, M., Sharma, E. A. (2016). Chronic Kidney Disease Detection by Analyzing Medical Datasets in Weka. International Journal of Computer Application, 6(4), 20-26

[13] M. K. J. Ms. Astha Ameta, "Data Mining Techniques for the Prediction of Kidney Diseases and Treatment: A Review," International Journal Of Engineering.

[14] A. Maurya, R. Wable, R. Shinde, S. John, R. Jadhav and R. Dakshayani, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-4, doi: 10.1109/ICNTE44896.2019.8946029.

[15] J. Snegha, V. Tharani, S. D. Preetha, R. Charanya and S. Bhavani, "Chronic Kidney Disease Prediction Using Data Mining," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-5, doi: 10.1109/icETITE47903.2020.482.

[16] Satukumati S.B., Satla S., Kogila R. (2019). Feature extraction techniques for chronic kidney disease identification, Ingenierie des Systemesd'Information, Vol. 24, No. 1, pp. 95-99. https://doi.org/10.18280/isi.240114.

[17] A. J. Aljaaf et al., "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, 2018, pp. 1-9, doi: 10.1109/CEC.2018.8477876.

[18] Siddheswar Tekale, Pranjal Shingavi, Sukanya Wandhekar, Ankit Chatorikar, "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm" International Journal of Advanced Research in Computer and Communication Engineering. Vol-7, issue-10, 2018.

[19] S. Pitchumani Angayarkanni. "Predictive Analysis of Chronic Kidney Disease Using Machine Learning Algorithm" International Journal of Recent Technology and Engineering (IJRTE). ISSN:2277-3878, vol-8, issue- 2,2019

[20] Deepika B, Rao VKR, Rampure DN, Prajwal P, Gowda DG, et al (2020) Early Prediction of Chronic Kidney Disease by using Machine Learning Techniques. Am J Comput Sci Eng Surv Vol. 8 No. 2:7.

[21] Marwa Almasoud, Tomas E Ward. "Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number o Predictors." International Journal o Advanced Science and Applications (IJACSA), vol. 10, No. 8, 2019.

[22] S. Pitchumani Angayarkanni. "Predictive Analysis of Chronic Kidney Disease Using Machine Learning Algorithm" International Journal of Recent Technology and Engineering (IJRTE). ISSN:2277-3878, vol-8, issue- 2,2019