LITERATURE REVIEW:

The reviews indicate that several studies have been conducted on chronic kidney disease prediction using machine-learning techniques. There are various parameters which play an important role in improving model performance like dataset size, quality of dataset and the time dataset collected.

The focus on chronic kidney disease prediction using machine learning models based on the dataset with a larger size aims to predict Three machine learning algorithms;

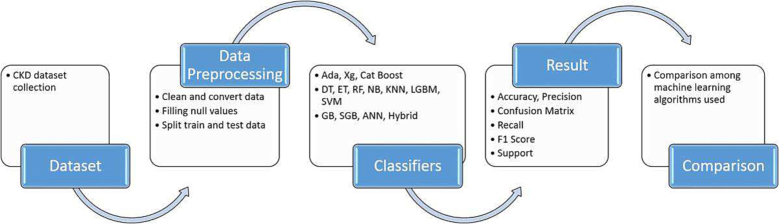
Random Forest, Support Vector Machine and Decision Tree .

The algorithms were selected based on their popularity in chronic kidney disease prediction and their performance of classification on previous research works

Such as **Priyanka et**  carried out chronic kidney disease prediction through naive bayes. They have tested using other algorithms such as KNN (K-Nearest Neighbor Algorithm), SVM (Support Vector Machines), Decision tree, and ANN (Artificial Neural Network) and they have got Naïve Bayes with better accuracy of 94.6% when compared to other algorithms.

Similarly, Random Forest has the highest voting results of the target output by each tree, Support Vector Machine

METHADOLODY



CONCLUSION:

This study presented a number of different machine learning algorithms with the intention of making a CKD diagnosis at an earlier stage. The models that are constructed using CKD patients are then trained and validated using the input parameters that were discussed earlier. Studies have been done on the associations between different factors so that the number of features can be cut down and redundant information eliminated. When applying a filter feature selection approach to the remaining attributes, it was discovered that hemoglobin, albumin, and specific gravity had the biggest impact when it comes to predicting CKD. This was the case after the method was used. This work presented a number of different machine learning algorithms with the intention of making a CKD diagnosis at an earlier stage. The original CKD dataset has been preprocessed first to validate the machine learning-based detection models. After that, the PCA has been performed to identify the most dominant features, thereby detecting CKD. The models that are constructed using CKD patients are then trained and validated using the input parameters that were discussed earlier. The accuracy of such algorithms was the primary criterion that was utilized in evaluating their overall performance.

RESULT:

The results of each classifier have been reviewed with a variety of criteria for evaluation, and the 10-fold cross-validation method has been utilized to validate the results against overfitting.

In addition, the layered cross-validation method has been utilized with the goal of fine-tuning the models’ respective parameter settings. The Google Colab web application, written in Python, is used to run the experiments, which are out using that programming language. This project has utilized DATAWORLD, which is open- source software for the machine learning library in Python, in several different ways. Accuracy, F1-score, precision, and recall are the evaluation metrics that were taken into consideration for this study.

The various values that are assigned to each model’s parameters result in the production of distinct sets of outputs

According to thE results, all of the models have excellent performance in terms of detecting CKD with accuracy of greater than 97% by utilizing hemoglobin, specific gravity, and albumin, sugar, blood glucose random, serum creatinine, potassium, packed cell volume, white and red blood cell count, and diabetes mellitus characteristics, shown in. By concentrating on precision and recall, Random forest tree 98%, with the exception of kNN, which indicates that most of the models were correct identifying the participants who did not have the disease or were healthy.