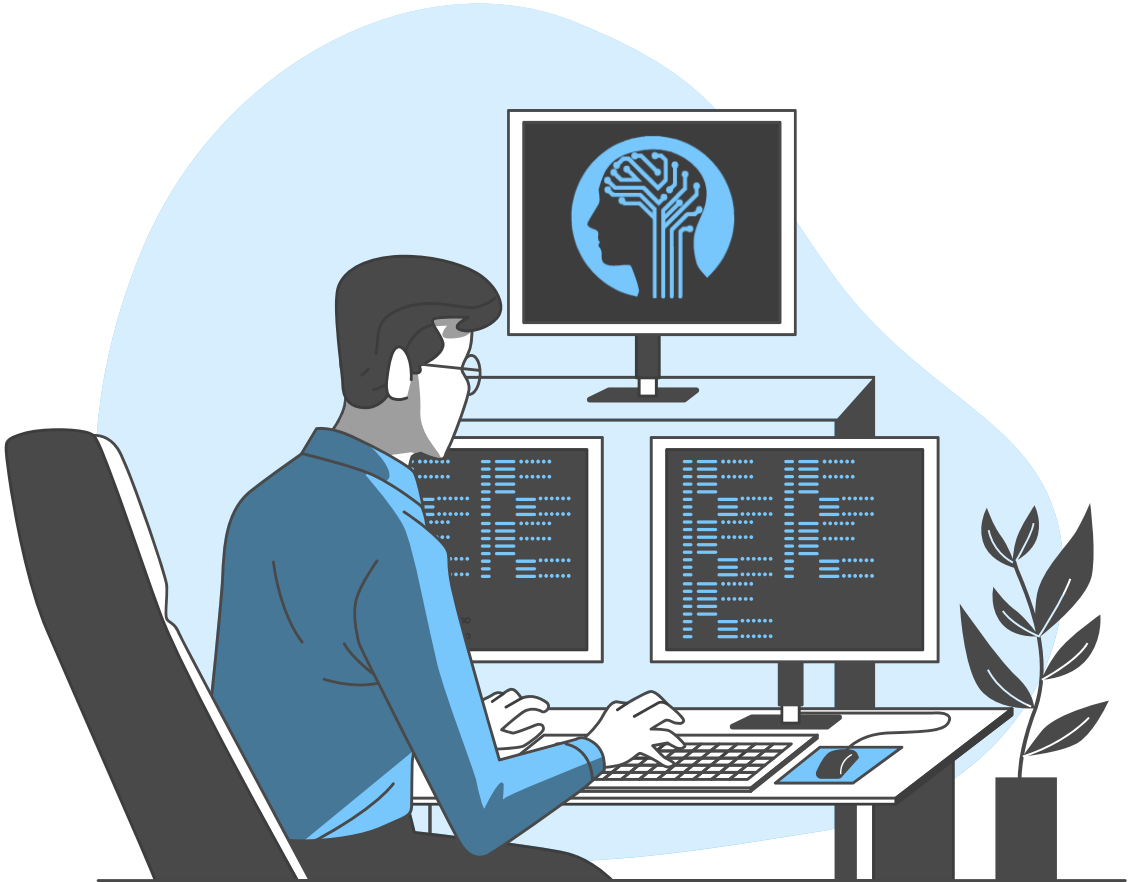


Machine Learning

Chronic Kidney Disease

Our Team

Team Members

Mostafa Hamdy Abo El-Ela

Nada Mohamed Salah

Nourhan Atef Mohamed

Yousef Ragab Abd El-Motaleb

Team Supervisors

Prof. Hala Zayed

Eng. AbdulWahab Almestekawy

Table of Contents

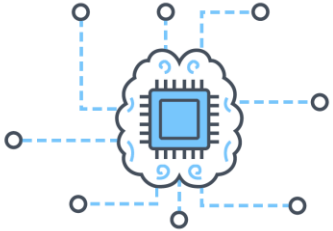
01

Introduction

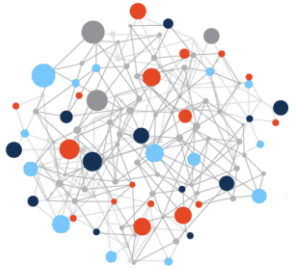
Machine Learning

Model Training

Train Data



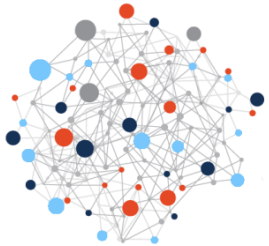
ML Algorithms



Predictive Model

Model Testing (Scoring)

New Data



Predictive Model

Predictions

02

Problem Statement

What is CKD?

Chronic kidney disease (CKD): is a long-term condition where the kidneys don't work as well as they should and cannot correctly filter blood.

CKD, in its early stages, has no symptoms; testing may be the only way to find out if the patient has kidney disease.

About 10% of the population worldwide suffers from (CKD), and millions die each year because they cannot get affordable treatment, with the number increasing in the elderly.

Our Goal

Early detection of CKD in its initial stages can help the patients get effective treatment and the sooner they know about having this disease, the sooner they can get treatment.

To predict positive CKD status and the stages of CKD machine learning can be used.

Machine Learning grabs a major part of artificial intelligence when it comes to doing predictions from previous data using classification and regression methods.

Dataset

These data has 400 instances which are a comparatively small number of samples with 25 attributes.

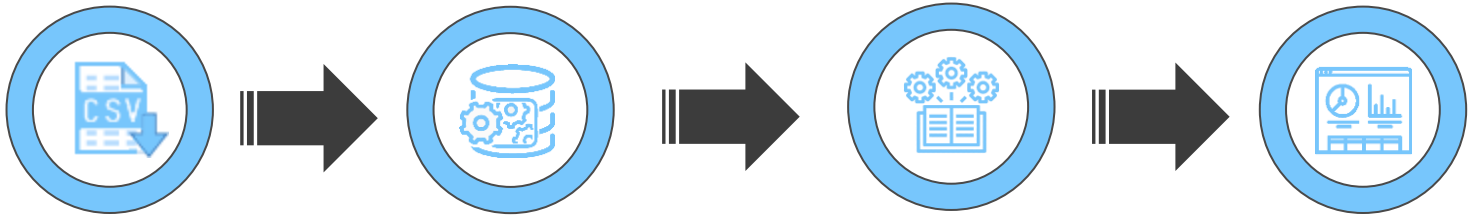
The data includes 250 records of patients with CKD and 150 records of persons without CKD.

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 and contains missing values.

03

Methodology

Methology



Read Data

- Import Dataset
- Replace “?” with “NaN” (Missing Values)

...

Data Preprocessing

- Clean Data (handle Missing Values)
- Split Data into: Input-Output
- Handle Categorical Data

...

Model Selection

- Split Data into: Train-Validation-Test
- Feature Scaling
- Classification & Testing
- Ensemble Learning

...

Evaluation Metrics

- Confusion Matrix
- Classification Report
- Correlation Coefficient
- Feature Importance
- Save Model and Scaler

...

04

Results & Discussion

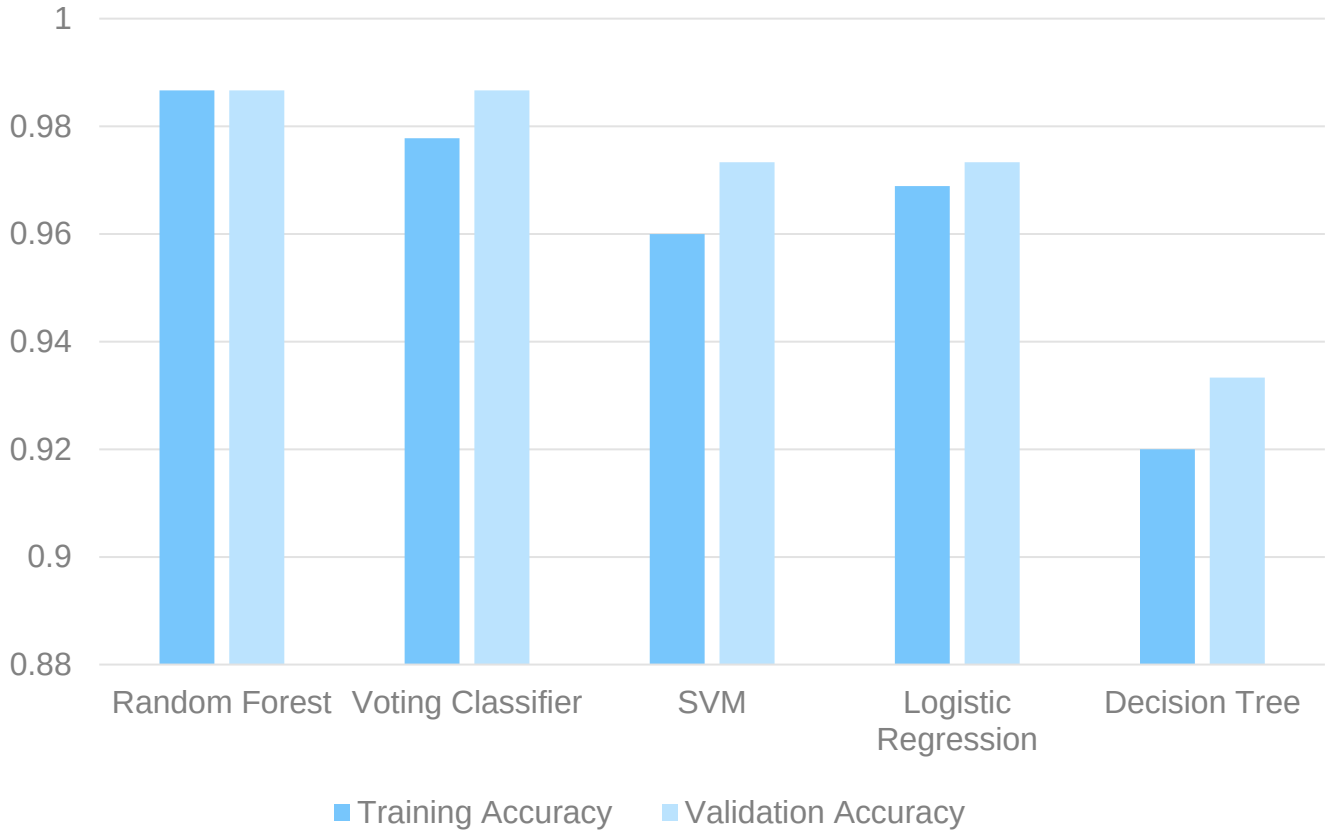
Models Accuracy

These are the results, which sorted descending according to validation accuracy.

As we have shown below, the best Model is Random Forest.

Index	Model	Training Accuracy	Validation Accuracy
0	Random Forest	0.986667	0.986667
1	Voting Classifer	0.977778	0.986667
2	SVM	0.960000	0.973333
3	Logistic Regression	0.968889	0.973333
4	Decision Tree	0.920000	0.933333

Models Accuracy Graph



Confusion Matrix

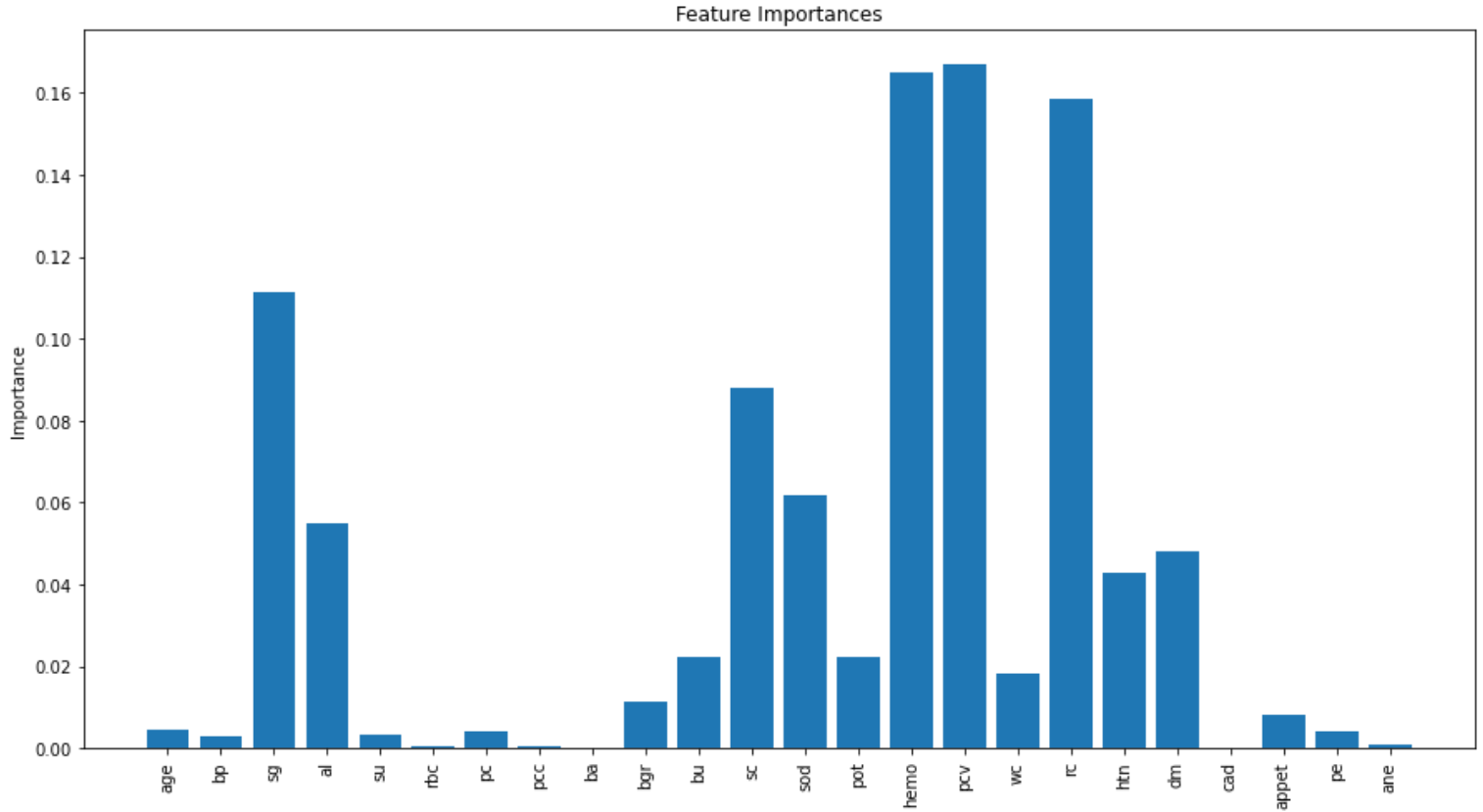
		Predict Values	
		CKD	Not CKD
Actual Values	CKD	TP = 62	FN = 0
	Not CKD	FP = 0	TN = 38

From this matrix, we can calculate: Accuracy, Precision, Recall, F1-score

Classification Report

Accuracy	$\frac{TP + TN}{TP + FN + FP + TN}$	1.00
Precision	$\frac{TP}{TP + FP}$	1.00
Recall	$\frac{TP}{TP + FN}$	1.00
F1-score	$\frac{2 \text{ x } (\text{Precision x Recall})}{\text{Precision} + \text{Recall}}$	1.00

Feature Importance



05

Conclusion

Conclusion

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 these machine learning classifiers: decision tree, logistic regression, SVM, random forest, and voting classifier on a small dataset of 400 records. After using a lot of models, we deduced that **Random forest** is the best model according to both training and validation accuracy.

A filter **feature selection** method has been applied to the remaining attributes and found that hemoglobin, albumin, Red blood cell count, Packed cell volume, and specific gravity have the most impact to predict the CKD.

References

https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

https://www.researchgate.net/publication/335698017_Detection_of_Chronic_Kidney_Disease_using_Machine_Learning_Algorithms_with_Least_Number_of_Predictors

https://ieeexplore.ieee.org/document/9185249

Thanks!