

Hope Artificial Intelligence



Classification Assignment

Problem Statement or Requirement:

A requirement from the Hospital, Management asked us to create a predictive model which will predict the Chronic Kidney Disease (CKD) based on the several parameters. The Client has provided the dataset of the same.

1.) Identify your problem statement

Domain Selection: - Machine Learning

Learning Selection: - Supervised Learning

Problem Identification:- Classification

2.) Tell basic info about the dataset (Total number of rows, columns)

Total number of Rows :- 399

Total number of Columns:- 25

3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)

One Hot Encoding method.

4.) Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.

Random Forest Classification

5.) All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.)

	precision	recall	f1-score	support
0	0.96	1.00	0.98	51
1	1.00	0.98	0.99	82
accuracy			0.98	133
macro avg	0.98	0.99	0.98	133
weighted avg	0.99	0.98	0.99	133

6.) Mention your final model, justify why u have chosen the same.

Precision - When the model predicts CKD, it's correct 100% of the time.

Recall - It detects 98% of actual CKD cases, very few false negatives.

F1-Score - Excellent balance between precision and recall.

Accuracy - Overall, 98% of predictions are correct.

The Random Forest model achieved **98% accuracy** on the test set, with a **recall of 98%** and **precision of 100%** for CKD cases. This means it correctly identifies nearly all patients with CKD while avoiding false positives. The model also performs well on non-CKD cases, with a perfect recall. These metrics demonstrate that Random Forest is not only accurate but also clinically reliable for early CKD detection.