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

The hospital management wants to build a predictive machine learning model that can predict a patient have Chronic Kidney Disease (CKD) or not. based on several medical parameters.

2.) Tell basic info about the dataset (Total number of rows, columns)
399 rows × 25 columns

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

Convert string to numerical for 'dm_no', 'dm_yes', 'cad_yes', 'appet_poor', 'appet_yes', 'pe_poor','pe_yes', 'ane_no', 'ane_yes' by using one hot encoding it is nominal

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.

```
from sklearn.model_selection import GridSearchCV # built model for grid search
from sklearn.ensemble import RandomForestClassifier
param_grid = {
    'criterion': ['gini', 'entropy', 'log_loss'],
    'max_depth': [None, 5, 10, 15],
    'n_estimators': [100, 200]
}

grid= GridSearchCV(RandomForestClassifier(),param_grid, refit=True, verbose=3, n_jobs=-1, scoring= 'f1_weighted')
grid.fit(x_train,y_train)
```

Compared with decision tree random forest as same accuracy and f1 score better balanced across precision and recall for both class. Slight higher roc auc.

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

a. **Decision tree: 1.ROC AUC: 0.9688888888888889**
2.ACCURACY: 0.97

b. Random forest:1. ROC AUC: 0.9967407407407407
2.ACCURACY:0.97

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

After evaluate I chose compared with decision tree random forest as same accuracy and f1 score better balanced across precision and recall for both class. Slight higher roc auc.

Random forest:1. ROC AUC: 0.9967407407407407
2.ACCURACY: 0.97