Classification Assignment

Description: The hospital management requested a predictive model to identify Chronic Kidney Disease (CKD) using patient data with several clinical parameters. The client provided a dataset for building and testing the model.

Disease Prediction Model Comparison

1. Decision Tree - **Best Parameters:** `{'criterion': 'gini', 'max_features': 'sqrt', 'splitter': 'random'}` - **Performance:** - Accuracy: 95% - Recall (Yes): 0.94 - F1-score (Yes): 0.96 - **Observations:** Decision Tree captures positive cases well but slightly underperforms overall compared to ensemble models.

2. Random Forest - **Best Parameters:** `{'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 10}` - **Performance:** - Accuracy: 99% - Recall (Yes): 0.99 - F1-score (Yes): 0.99 - **Observations:** High accuracy and balanced precision/recall. Reduces overfitting and effectively captures complex patterns.

3. Support Vector Machine (SVM) - **Best Parameters:** `{'C': 10, 'gamma': 'auto', 'kernel': 'linear'}` - **Performance:** - Accuracy: 98% - Recall (Yes): 0.98 - F1-score (Yes): 0.99 - **Observations:** Excellent at distinguishing between classes. Slightly lower recall compared to Random Forest.

4. Naive Bayes - **Best Parameters:** `{'var_smoothing': 1e-09}` - **Performance:** - Accuracy: 98% - Recall (Yes): 0.96 - F1-score (Yes): 0.98 - **Observations:** Works well with continuous features but assumes feature independence, which might not capture all correlations.

Conclusion

For disease prediction, **Random Forest** is the preferred model. It delivers the highest accuracy (99%) while maintaining excellent recall (0.99) for positive cases. This ensures minimal false negatives, which is critical in medical diagnostics. While SVM and Naive Bayes also show strong performance, Random Forest offers superior robustness and generalization.

Recommendation

Deploy **Random Forest** for predicting CKD in real-world settings to maximize patient safety and ensure accurate identification of positive cases without compromising overall performance.