## Decision Trees

The results indicate perfect performance (1.0000) for both the test set and 10-fold cross-validation, which suggests the model is likely overfitting. This means the Decision Tree has learned the training data too well, including noise or patterns that don't generalize to unseen data. The dataset may be simple, clean, or perfectly separable, which allowed the model to achieve these ideal results.

Experimenting with various decision tree parameters that control the size of the tree:

| Best Parameters | {'criterion': 'gini', 'max\_depth': 3, 'min\_samples\_split': 2, 'min\_samples\_leaf': 1} |
| --- | --- |
| Best Cross-Validation Accuracy | 1.0000 (100%) |
| Test Accuracy | 1.0000 (100%) |
| Confusion Matrix | | [[18079, 0, 0, 0], | | --- |  |  | [0, 5700, 0, 0], | | --- | --- |  |  | [0, 0, 55367, 0], | | --- | --- |  |  | [0, 0, 0, 50725]] | | --- | --- | |

Classification Report (Test Set):

precision recall f1-score support

1 1.00 1.00 1.00 18079

2 1.00 1.00 1.00 5700

3 1.00 1.00 1.00 55367

4 1.00 1.00 1.00 50725

accuracy 1.00 129871

macro avg 1.00 1.00 1.00 129871

weighted avg 1.00 1.00 1.00 129871

### **Explanation:**

* **Confusion Matrix**: Perfect classification with no false positives or false negatives across all classes.
* **Classification Report**: Precision, recall, and F1-score for all classes (1, 2, 3, 4) are 1.00, indicating a flawless classification for each class.
* **ROC AUC**: The ROC AUC score of 1.0000 confirms perfect model performance in distinguishing between the classes.

**Tree 1:**

Answer to the Questions:

1. Does the decision tree generalize well to new data?

Yes, the decision tree generalizes very well to new data. The test accuracy is 1.0000, which means the model is correctly classifying all instances in the test set. Additionally, the confusion matrix shows that there are no misclassifications, with all predicted values matching the actual values for each class.

2.How do you tell?

- Test Accuracy: The model achieved 100% accuracy on the test set, indicating perfect performance on unseen data. This suggests that the model has learned the underlying patterns well and is not overfitting the training data.

- Confusion Matrix: The confusion matrix confirms that the model correctly predicted every instance for all classes (1, 2, 3, and 4), with no false positives or false negatives, further supporting the conclusion that the model generalizes well.

- Classification Report: The precision, recall, and F1-score for all classes are 1.00, showing that the model has perfect performance across all metrics. This means the model is not only correctly classifying each instance but is also highly reliable and consistent across all classes.

- ROC AUC: The ROC AUC score of 1.0000 indicates that the model distinguishes between different classes perfectly, which is a strong indicator of its generalization ability.

Conclusion:

The decision tree classifier, with the best parameters, performs exceptionally well on the test data, with perfect accuracy and no misclassifications. This indicates that the model generalizes effectively to new data and is likely not overfitting the training set.

**Tree 2:**

Key Findings:

Training Accuracy: 1.0000

Test Accuracy: 1.0000

These results indicate that the decision tree classifier performs perfectly on both the training and test sets.

The decision tree generalizes exceptionally well to new data. The accuracy, confusion matrix, and classification report all show perfect performance on the test set, which indicates that the model has not overfitted to the training data.

How do you tell? The perfect performance on the test set, which is representative of unseen data, suggests the model is robust and generalizes well. There is no discrepancy between the performance on the training set and the test set, further confirming that the model is not overfitting.

In conclusion, the decision tree classifier is performing optimally and generalizing well to the new data after moving 30% of the instances to the testing set.

**Tree 3:**

Results Analysis Training Accuracy: 1.0000 Test Accuracy: 1.0000 These results show that the classifier has perfect accuracy on both the training and test sets, meaning it is able to correctly classify every instance in both datasets.