

# FAST NATIONAL UNIVERSITY CFD CAMPUS

|  |  |
| --- | --- |
| **Name:** | **Mariam Fatima**  **Muskan Ghani** |
| **Roll No:** | **22F3168**  **22F3841** |
| **Assignment:** | **2** |
| **Instructor:** | **Dr. Hashim Yaseen** |
| **Course** | **Machine Learning** |

# Decision Tree Classifiers (ID3 and CART)

## Implementation and Results:

### ID3 Decision Tree:

**Training Accuracy:** 97.5%

**Testing Accuracy:** 93.33%

### CART Decision Tree:

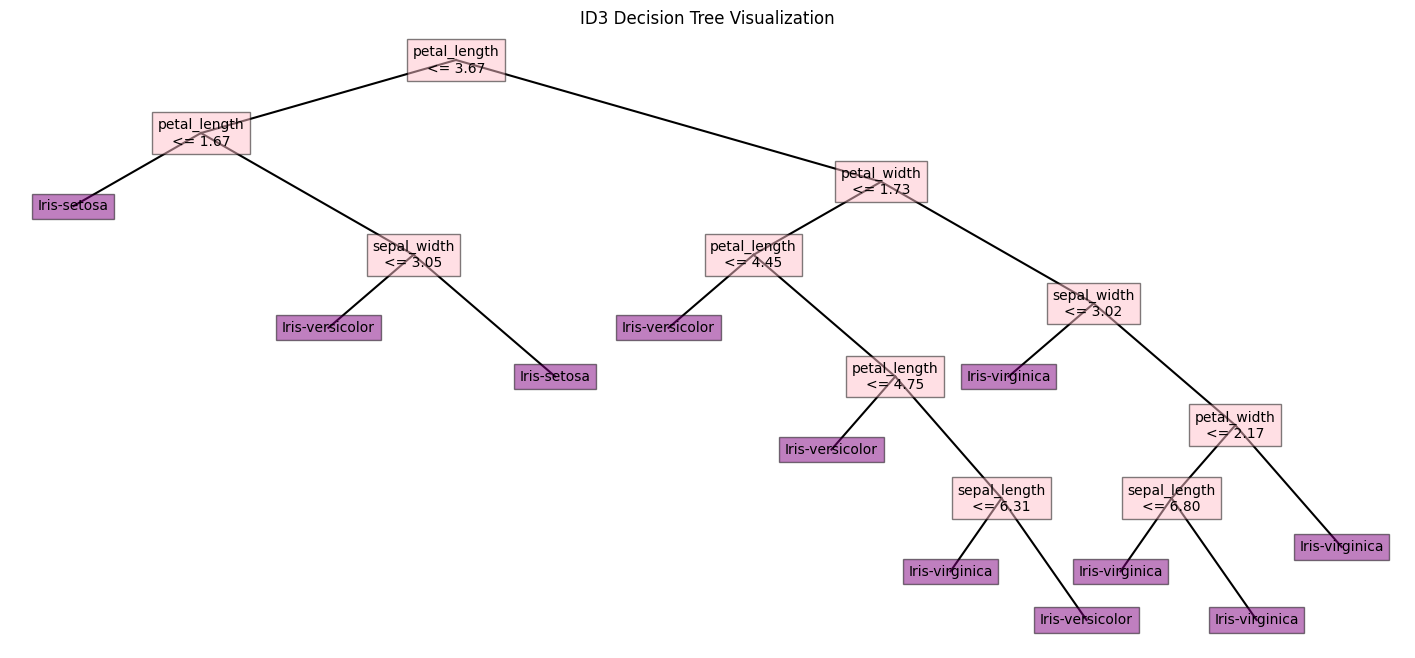
**Training Accuracy:** 100.0%

**Testing Accuracy:** 93.33%

Both the ID3 and CART decision trees performed well on the Iris dataset, achieving high training and testing accuracies. CART, which uses the Gini index for splitting, achieved perfect training accuracy, indicating it overfits the training data. However, both models have the same testing accuracy, suggesting similar generalization performance.

## Visualization:

The tree structures for both ID3 and CART are visualized. These diagrams illustrate the hierarchical decision-making process, where the ID3 and CART trees split based on the features with the highest information gain (ID3) or Gini index reduction (CART).



A diagram of a cat

Description automatically generated

## Evaluation:

Overall, both ID3 and CART decision trees show robust classification accuracy on the Iris dataset, with minimal variance between training and testing performance. This indicates low overfitting and good generalization capabilities for both algorithms.

# Random Forest Classifier

## Implementation and Results:

**Optimal Number of Trees:** 3

**Training Accuracy with 3 Trees:** 98%

**Testing Accuracy with 3 Trees:** 100%

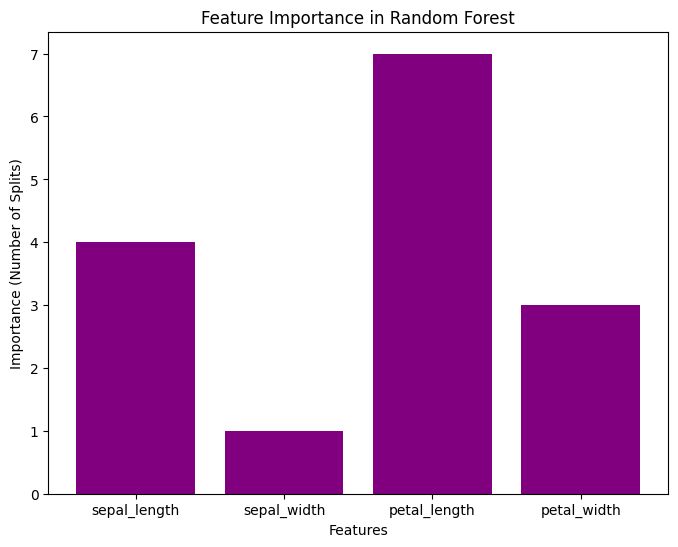
The Random Forest model, trained with an optimal number of 3 trees, achieved near perfect accuracy on both training and testing datasets. This improvement in accuracy can be attributed to the ensemble learning approach, where multiple trees are aggregated to make more accurate predictions. The model's high testing accuracy reflects its excellent generalization capability.

## Visualization and Validation:

The choice of 3 trees was validated as optimal due to the observed stabilization in accuracy. Visualizations of the accuracy with different tree counts showed that after reaching 3 trees, the marginal benefit in terms of accuracy diminished. Therefore, 3 trees strike a balance between performance and computational efficiency.

A graph with orange and blue lines

Description automatically generated



# Naïve Bayes Classifier

## Implementation and Results:

**Testing Accuracy:** 93.33%

The Naïve Bayes classifier, a probabilistic approach, yielded a test accuracy of **93.33%,** matching the testing performance of both ID3 and CART. This classifier assumes feature **independence**, which simplifies computation but can affect accuracy if features are correlated. However, Naïve Bayes performs competitively on the Iris dataset, which indicates it is well suited for this data.

A diagram of a data set

Description automatically generated with medium confidence

# Comparative Analysis of Classifiers

## Performance Comparison:

All three classifiers—ID3, CART, and Naïve Bayes, achieved similar testing accuracy (93.33%). However, their training accuracies and the method of construction varied:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ID3 | CART | RANDOM FOREST | NAÏVE BAYES |
| Optimal no. of trees | - | - | 3 | - |
| Training accuracy | 97.50% | 100% | 98% | - |
| Testing accuracy | 93.33% | 93.33% | 100% | 93.33% |

**ID3 and CART:** CART’s perfect training accuracy suggests a complete fit to the training data, while ID3 remains highly accurate yet slightly less overfitted.

**Random Forest:** Demonstrated the **highest testing accuracy of 100%,** showing the advantages of ensemble methods in enhancing prediction accuracy and robustness.

**Naïve Bayes:** Competitive accuracy with a simpler probabilistic approach, indicating effectiveness for this dataset despite its independence assumption.

## Visualization of Results:

The Random Forest outperformed with a slight edge, indicating that ensemble methods can yield significant accuracy benefits over single decision trees. Both decision trees and Naïve Bayes provided consistent results, underscoring their reliability for classification tasks.