Machine Learning – Task 3

Submitters

- Ron Ashwal 208573139
- Zohar Simhon 211871868
- Git: https://github.com/ZoharSimhon/Decision-Tree

Project Overview

This project implements and compares two decision tree algorithms for binary classification problems:

1. Optimized Brute-Force Method
2. Binary Entropy-Based Method

The implementation is designed to analyze the performance of these algorithms across various tree depths (k) and visualize the resulting decision trees.

Methodology

Both algorithms construct decision trees by recursively splitting the data based on features.

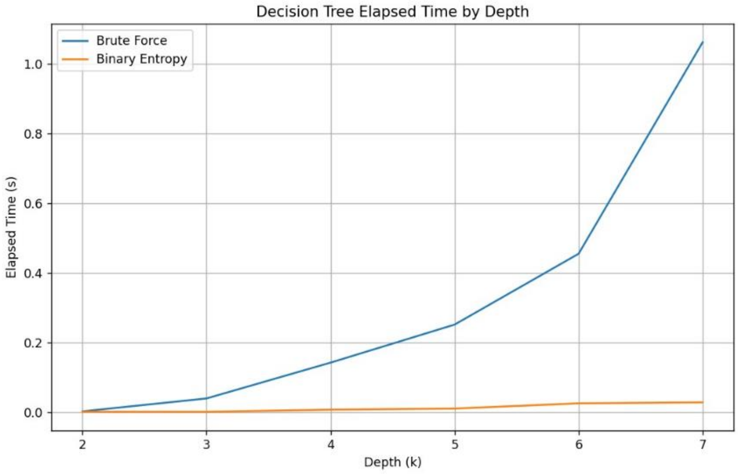
The key differences are:

- **Brute-Force Method**: Exhaustively searches for the best feature to split on at each node, optimized using dynamic programming and caching.

- **Binary Entropy Method**: Uses the concept of binary entropy to determine the best feature for splitting, aiming to maximize information gain.

Results

- **Time results:**

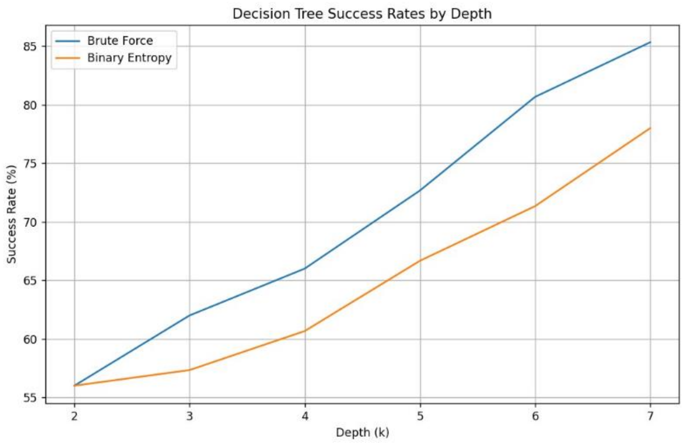


The graph above shows the execution times for both methods across different tree depths (k).

<u>Key observations</u>:

  - ○ The binary entropy method consistently executes faster than the brute-force method.
  - ○ As k increases, the execution time for both methods increases, with the brute-force method showing a more pronounced increase.

- **Success Rate Comparison:**



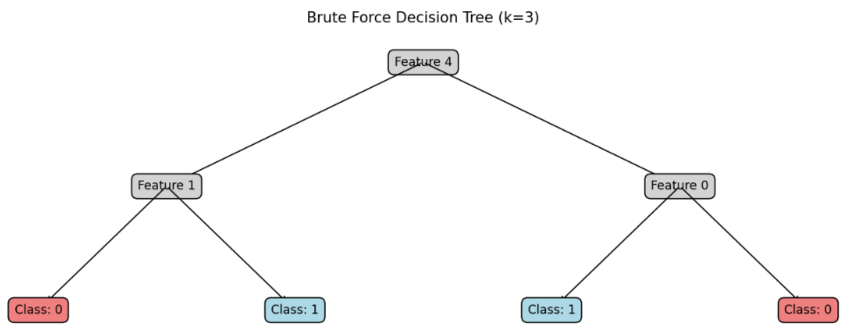
This graph compares the success rates of both methods for various tree depths.

<u>Notable findings</u>:

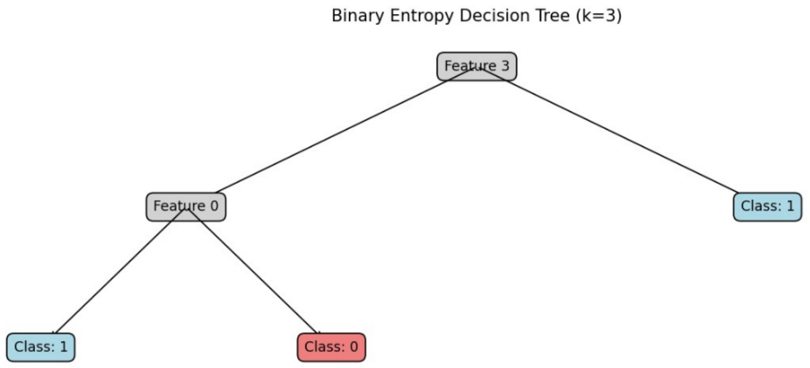
- ○ The brute-force method consistently achieves higher success rates compared to the binary entropy method.
- ○ -Both methods show improved success rates as k increases, with diminishing returns at higher depths.

Decision Tree Structures

- **Brute-Force Method Tree (k=3)**
  - Success Rate: 62.00%
  - Execution Time: 0.0412 second



- **Binary Entropy Method Tree (k=3)**
  - Success Rate: 57.33 %
  - Execution Time: 0.0106 second



These visualizations demonstrate the structural differences between the trees generated by each method:

- The brute-force method tends to create more balanced trees with potentially better feature selections at each node.

- The binary entropy method may produce less balanced trees but generates them more quickly.

Key Findings

1. **Accuracy**: The brute-force method consistently outperforms the binary entropy method in terms of accuracy for any given depth (k).

2. **Execution Time**: The binary entropy method is generally faster than the brute-force method for any given depth (k).

3. **Depth Impact**: Increasing the depth (k) of the decision trees typically leads to improved results for both methods, with a trade-off between accuracy and computational cost.

Conclusions

This project demonstrates the trade-offs between accuracy and computational efficiency in decision tree algorithms. The brute-force method offers superior accuracy but at the cost of longer execution times, especially for larger datasets or greater tree depths. The binary entropy method provides a faster alternative with slightly lower accuracy, making it potentially more suitable for large-scale applications or when quick results are needed.