**Warsaw University of Technology**
Faculty of Electronics and Information Technology
Specialization: Computer Science
Course: Introduction to Artificial Intelligence

# PROJECT DOCUMENTATION

## ID3 Algorithm

**Author:**
Monika Jung

**Student ID:**
331384

# 1 Task Description

The goal of this project was to implement a decision tree classifier using the ID3 algorithm with a restriction on the maximum depth of the tree. The classifier was evaluated on the Tic-Tac-Toe Endgame dataset to measure its accuracy and produce a confusion matrix.

# 2 Dataset

**Source:** UCI Machine Learning Repository **Name:** Tic-Tac-Toe Endgame **Type:** Multivariate, Categorical, Binary Classification **Number of instances:** 958 **Features:** 9 board positions (values: `x, o, b`) **Target:** Class (`positive` = X wins, `negative` = otherwise) **Missing values:** No

# 3 Implementation

The ID3 algorithm was implemented from scratch. The main components included:

- Entropy and Information Gain calculation
- Recursive tree construction with maximum depth limit
- Predict function for classification

The dataset was split into three parts:

- 60
- 20
- 20

# 4 Evaluation Results

## Validation Set

- Accuracy: 82.29
- Confusion Matrix:

```
[[ 45  14]
[ 20 113]]
```

- Classification Report:

```
precision    recall  f1-score   support

       0       0.69      0.76      0.73        59
       1       0.89      0.85      0.87       133

accuracy                           0.82       192

macro avg       0.79      0.81      0.80       192
weighted avg       0.83      0.82      0.83       192
```

## Test Set

- Accuracy: 81.25
- Confusion Matrix:

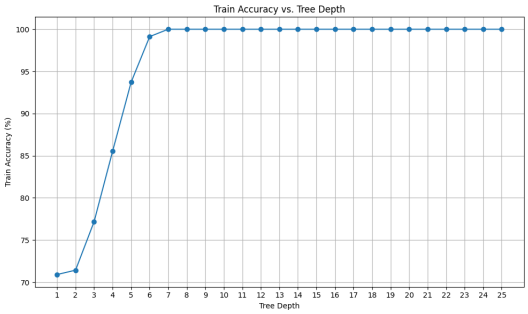
```
[[ 43  19]
[ 17 113]]
```

- Classification Report:

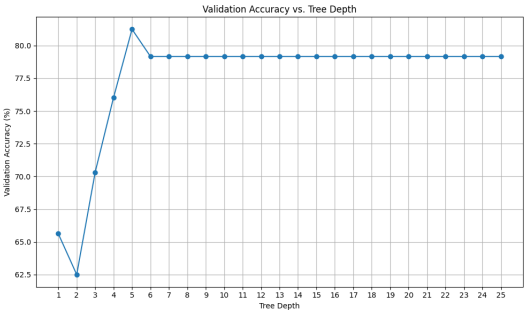
```
precision    recall  f1-score   support

       0       0.72      0.69      0.70        62
       1       0.86      0.87      0.86       130

accuracy                           0.81       192

macro avg       0.79      0.78      0.78       192
weighted avg       0.81      0.81      0.81       192
```

# 5 Visualization

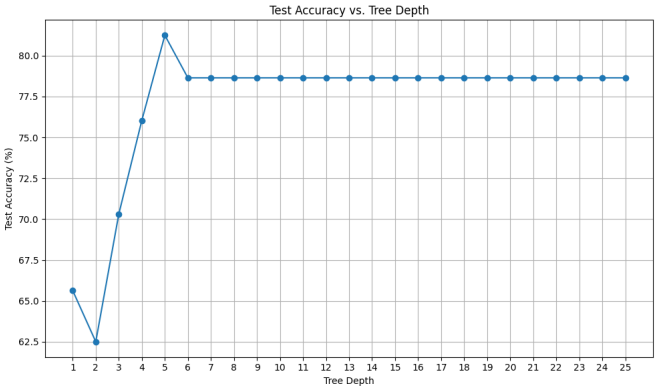
## Accuracy vs. Tree Depth




(a) Train Accuracy




(b) Validation Accuracy




(c) Test Accuracy

Figure 1: Accuracy trends for different tree depths.

## Interpretation

The validation and test accuracy curves both peak around depth 4–5, after which they stabilize. This suggests that the model learns all generalizable patterns by this depth. Increasing the depth beyond this point leads to no performance gain on unseen data.

The training accuracy, on the other hand, continues to increase and quickly reaches 100% around depth 6.

The optimal depth is therefore where validation accuracy is maximized—in this case, around 4 or 5.

### Why Accuracy Levels Do Not Rise After Depth 6

The dataset consists of finite and relatively simple board configurations with categorical features (`x, o, b`). The ID3 algorithm splits data based on the most informative attributes until:

- all instances in a node have the same class,
- no attributes are left, or
- the maximum depth is reached.

At around depth 5–6, the most useful decision rules have already been learned. Additional splits do not improving generalization, as confirmed by the flattening accuracy curve.

# 6 Depth more than 8 and the Absence of Overfitting

When analyzing the results, I initially expected to observe accuracy degradation at higher maximum depths due to overfitting. However, the experimental data revealed no such deterioration.

The debug output reveals why overfitting doesn't occur beyond depth 7: I added a simple print function that prints the reason why algorithm was terminated.

```
Current depth: 7, all labels are the same.
(Reapated everytime when max depth >= 8)
```

- **In general the ID3 algorithm stops when**:
  1. All samples in a node have identical labels (your most frequent case)
  2. No attributes remain for splitting
  3. Maximum depth is reached (but rarely triggers)
- **In this case**:
  - Decision paths terminate by depth 7 due to complete label purity
  - Additional depth levels remain unused because the tree cannot grow further