Draft Blog Entry: Translating ID3 into Modern C++

Author: James (Pate) Williams, Jr.
Additional narrative assistance: Microsoft Copilot (credited explicitly below)

Introduction

This project represents the latest chapter in a long-running personal exploration of machine-learning algorithms. My first implementation of Quinlan's ID3 algorithm dates back to the late 1990s, when I wrote a C version inspired by Tom M. Mitchell's clear and mathematically grounded presentation in *Machine Learning*. I later produced a C# variant around 2015–2016, refining the structure and improving the clarity of the entropy and information-gain calculations.

In early 2026, I returned to ID3 with the goal of producing a clean, modern C++ implementation — one that is deterministic, modular, and faithful to Mitchell's formulation. This post documents that translation and offers readers a look at the architecture, the dataset, and the resulting decision tree.

Project Structure

My C++ implementation consists of **four header files** and **five source files**, all written by hand. I do not include compiler-generated files or IDE scaffolding. The goal was to keep the project compact, readable, and easy to study.

The codebase is organized around the following components:

- **Dataset parsing and representation**
- **Entropy and information-gain computation**
- **Recursive tree construction**
- **Deterministic tie-breaking (first-come-first-served)**
- **Tree traversal and classification output**

This structure mirrors the conceptual flow of Mitchell's pseudocode while taking advantage of modern C++ features for clarity and safety.

The Luger Credit-Risk Dataset

To test the implementation, I used the credit-risk example from Luger and Stubblefield's *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. This dataset is small but rich enough to illustrate how ID3 behaves when faced with:

- categorical attributes
- “Unknown” values
- mixed distributions of class labels

My program’s verbose output (included in the PDF version of this post) shows the entropy of each branch, the information gain for each attribute, and the resulting splits.

One interesting observation:
Luger’s printed tree uses **Credit History** as the root attribute, while my implementation selects **Income**. This is not an error — it reflects differences in how textbooks handle “Unknown” values, rounding, and tie-breaking. My implementation follows Mitchell’s formulation precisely, and the resulting tree is mathematically correct.

Determinism and Tie-Breaking

In my evolutionary algorithms and genetic-algorithm experiments, I use random tie-breaking to encourage exploration.
For ID3, however, I use **first-come-first-served** tie resolution. This ensures:

- reproducibility
- cross-language consistency
- archival clarity

The same dataset will always produce the same tree, which is essential for long-term comparison across my C, C#, and C++ versions.

Verbose Output and Snippets

The PDF version of this post includes Snipping Tool captures of the program’s console output, showing:

- the full training set
- entropy calculations
- information-gain tables
- the recursive structure of the resulting decision tree

These images provide a clear view of the algorithm’s internal reasoning.

Availability of the Source Code

Readers who wish to study the implementation in detail may request the full Visual Studio project (headers and CPP files) in zipped format.
Please contact me via email, jamespate@mac.com and I will be happy to share the archive.

Closing Thoughts

This C++ translation completes a three-language lineage of ID3 implementations spanning nearly three decades. It also sets the stage for future work, including a C4.5 implementation later this year. For now, I'm pleased to share this clean, modern version of ID3 with anyone interested in machine-learning algorithms, archival software craftsmanship, or the evolution of my own technical projects.

Credit Risk Domain Example

```
None       $0-$15k    High      Bad       High
None       $15k-$35k  High      Unknown   High
None       $15k-$35k  Low       Unknown   Moderate
None       $0-$15k    Low       Unknown   High
None       Over $35k  Low       Unknown   Low
Adequate   Over $35k  Low       Unknown   Low
None       $0-$15k    Low       Bad       High
Adequate   Over $35k  Low       Bad       Moderate
None       $15k-$35k  Low       Good      Low
Adequate   Over $35k  High      Good      Low
None       $0-$15k    High      Good      High
None       $15k-$35k  High      Good      Moderate
None       Over $35k  High      Good      Low
None       $15k-$35k  High      Bad       High

Collateral  2 -0.196778 -1.127792 +0.206050
Income      3 -0.000000 -0.543546 -0.257831 +0.729242
Debt        2 -0.689392 -0.778328 +0.062899
History     3 -0.489625 -0.231794 -0.543546 +0.265654

Income

None       $0-$15k    High      Bad       High
None       $0-$15k    Low       Unknown   High
None       $0-$15k    Low       Bad       High
None       $0-$15k    High      Good      High

High Risk leaf node

None       $15k-$35k  High      Unknown   High
None       $15k-$35k  Low       Unknown   Moderate
None       $15k-$35k  Low       Good      Low
None       $15k-$35k  High      Good      Moderate
None       $15k-$35k  High      Bad       High

Collateral  2 -0.000000 -1.521928 +0.000000
Debt        2 -0.550978 -0.400000 +0.570951
History     3 -0.400000 -0.000000 -0.400000 +0.721928

History

None       $15k-$35k  Low       Good      Low
None       $15k-$35k  High      Good      Moderate

Collateral  2 -0.000000 -1.000000 +0.000000
Debt        2 -0.000000 -0.000000 +1.000000

Debt

None       $15k-$35k  High      Good      Moderate

Moderate Risk leaf node

None       $15k-$35k  Low       Good      Low

Low Risk leaf node

None       $15k-$35k  High      Bad       High

High Risk leaf node

None       $15k-$35k  High      Unknown   High
None       $15k-$35k  Low       Unknown   Moderate

Collateral  2 -0.000000 -1.000000 +0.000000
Debt        2 -0.000000 -0.000000 +1.000000

Debt

None       $15k-$35k  High      Unknown   High

High Risk leaf node

None       $15k-$35k  Low       Unknown   Moderate

Moderate Risk leaf node

None       Over $35k  Low       Unknown   Low
```

```
Adequate   Over $35k  Low       Unknown   Low
Adequate   Over $35k  Low       Bad       Moderate
Adequate   Over $35k  High      Good      Low
None       Over $35k  High      Good      Low

Collateral  2 -0.550978 -0.000000 +0.170951
Debt        2 -0.000000 -0.550978 +0.170951
History     3 -0.000000 -0.000000 -0.000000 +0.721928

History

Adequate   Over $35k  High      Good      Low
None       Over $35k  High      Good      Low

Low Risk leaf node

Adequate   Over $35k  Low       Bad       Moderate

Moderate Risk leaf node

None       Over $35k  Low       Unknown   Low
Adequate   Over $35k  Low       Unknown   Low

Low Risk leaf node
```

Play Tennis Example (from ***Machine Learning*** by Tom M. Mitchell)

```
Sunny    Hot  High   Weak   No
Sunny    Hot  High   Strong No
Overcast Hot  High   Weak   Yes
Rain     Mild High   Weak   Yes
Rain     Cool Normal Weak   Yes
Rain     Cool Normal Strong No
Overcast Cool Normal Strong Yes
Sunny    Mild High   Weak   No
Sunny    Cool Normal Weak   Yes
Rain     Mild Normal Weak   Yes
Sunny    Mild Normal Strong Yes
Overcast Mild High   Strong Yes
Overcast Hot  Normal Weak   Yes
Rain     Mild High   Strong No

Outlook     3 -0.000000 -0.346768 -0.346768 +0.246750
Temperature 3 -0.231794 -0.285714 -0.393555 +0.029223
Humidity    2 -0.492614 -0.295836 +0.151836
Wind        2 -0.428571 -0.463587 +0.048127
```

```
Outlook

Overcast Hot  High   Weak   Yes
Overcast Cool Normal Strong Yes
Overcast Mild High   Strong Yes
Overcast Hot  Normal Weak   Yes

Yes leaf node

Rain     Mild High   Weak   Yes
Rain     Cool Normal Weak   Yes
Rain     Cool Normal Strong No
Rain     Mild Normal Weak   Yes
Rain     Mild High   Strong No

Temperature 3 -0.400000 -0.000000 -0.550978 +0.019973
Humidity    2 -0.400000 -0.550978 +0.019973
Wind        2 -0.000000 -0.000000 +0.970951

Wind

Rain     Cool Normal Strong No
Rain     Mild High   Strong No

No leaf node

Rain     Mild High   Weak   Yes
Rain     Cool Normal Weak   Yes
Rain     Mild Normal Weak   Yes

Yes leaf node

Sunny    Hot  High   Weak   No
Sunny    Hot  High   Strong No
Sunny    Mild High   Weak   No
Sunny    Cool Normal Weak   Yes
Sunny    Mild Normal Strong Yes

Temperature 3 -0.000000 -0.000000 -0.400000 +0.570951
Humidity    2 -0.000000 -0.000000 +0.970951
Wind        2 -0.400000 -0.550978 +0.019973

Humidity

Sunny    Hot  High   Weak   No
Sunny    Hot  High   Strong No
Sunny    Mild High   Weak   No

No leaf node

Sunny    Cool Normal Weak   Yes
Sunny    Mild Normal Strong Yes

Yes leaf node
```

Will Wait Restaurant Domain Example (Russell and Norvig)

```
Yes No  No  Yes Some $$$ No  Yes French  00 - 10 Yes
Yes No  No  Yes Full $   No  No  Thai    30 - 60 No
No  Yes No  No  Some $   No  No  Burger  00 - 10 Yes
Yes No  Yes Yes Full $   No  No  Thai    10 - 30 Yes
Yes No  Yes No  Full $$$ No  Yes French  GT   60 No
No  Yes No  Yes Some $$  Yes Yes Italian 00 - 10 Yes
No  Yes No  No  None $   Yes No  Burger  00 - 10 No
No  No  No  Yes Some $$  Yes Yes Thai    00 - 10 Yes
No  Yes Yes No  Full $   Yes No  Burger  GT   60 No
Yes Yes Yes Yes Full $$$ No  Yes Italian 10 - 30 No
No  No  No  No  None $   No  No  Thai    00 - 10 No
Yes Yes Yes Yes Full $   No  No  Burger  30 - 60 Yes

Alternate   2 -0.500000 -0.500000 +0.000000
Bar         2 -0.500000 -0.500000 +0.000000
Fridays     2 -0.574716 -0.404563 +0.020721
Hungry      2 -0.300803 -0.503487 +0.195710
Patrons     3 -0.000000 -0.000000 -0.459148 +0.540852
Price       3 -0.574716 -0.000000 -0.229574 +0.195710
Rain        2 -0.666667 -0.333333 +0.000000
Reservation 2 -0.574716 -0.404563 +0.020721
Type        4 -0.166667 -0.166667 -0.333333 -0.333333 +0.000000
Estimate    4 -0.000000 -0.166667 -0.166667 -0.459148 +0.207519

Patrons

No  Yes No  No  None $   Yes No  Burger  00 - 10 No
No  No  No  No  None $   No  No  Thai    00 - 10 No

No leaf node

Yes No  No  Yes Some $$$ No  Yes French  00 - 10 Yes
No  Yes No  No  Some $   No  No  Burger  00 - 10 Yes
No  Yes No  Yes Some $$  Yes Yes Italian 00 - 10 Yes
No  No  No  Yes Some $$  Yes Yes Thai    00 - 10 Yes
```

```
Yes leaf node

Yes No  No  Yes Full $   No  No  Thai    30 - 60 No
Yes No  Yes Yes Full $   No  No  Thai    10 - 30 Yes
Yes No  Yes No  Full $$$ No  Yes French  GT   60 No
No  Yes Yes No  Full $   Yes No  Burger  GT   60 No
Yes Yes Yes Yes Full $$$ No  Yes Italian 10 - 30 No
Yes Yes Yes Yes Full $   No  No  Burger  30 - 60 Yes

Alternate   2 -0.000000 -0.809125 +0.109170
Bar         2 -0.459148 -0.459148 +0.000000
Fridays     2 -0.000000 -0.809125 +0.109170
Hungry      2 -0.000000 -0.666667 +0.251629
Price       3 -0.666667 -0.000000 -0.000000 +0.251629
Rain        2 -0.809125 -0.000000 +0.109170
Reservation 2 -0.666667 -0.000000 +0.251629
Type        4 -0.000000 -0.000000 -0.333333 -0.333333 +0.251629
Estimate    4 -0.000000 -0.333333 -0.333333 -0.000000 +0.251629

Hungry

Yes No  Yes No  Full $$$ No  Yes French  GT   60 No
No  Yes Yes No  Full $   Yes No  Burger  GT   60 No

No leaf node

Yes No  No  Yes Full $   No  No  Thai    30 - 60 No
Yes No  Yes Yes Full $   No  No  Thai    10 - 30 Yes
Yes Yes Yes Yes Full $$$ No  Yes Italian 10 - 30 No
Yes Yes Yes Yes Full $   No  No  Burger  30 - 60 Yes

Alternate   2 -0.000000 -1.000000 +0.000000
Bar         2 -0.500000 -0.500000 +0.000000
Fridays     2 -0.000000 -0.688722 +0.311278
Price       3 -0.688722 -0.000000 -0.000000 +0.311278
Rain        2 -1.000000 -0.000000 +0.000000
Reservation 2 -0.688722 -0.000000 +0.311278
Type        4 -0.000000 -0.000000 -0.500000 -0.000000 +0.500000
Estimate    4 -0.000000 -0.500000 -0.500000 -0.000000 +0.000000

Type

No  No  No  No  None $   No  No  French  GT   60 No

Yes leaf node
Yes Yes Yes Yes Full $$$ No  Yes Italian 10 - 30 No

No leaf node

Yes No  No  Yes Full $   No  No  Thai    30 - 60 No
Yes No  Yes Yes Full $   No  No  Thai    10 - 30 Yes

Alternate   2 -0.000000 -1.000000 +0.000000
Bar         2 -1.000000 -0.000000 +0.000000
Fridays     2 -0.000000 -0.000000 +1.000000
Price       3 -1.000000 -0.000000 -0.000000 +0.000000
Rain        2 -1.000000 -0.000000 +0.000000
Reservation 2 -1.000000 -0.000000 +0.000000
Estimate    4 -0.000000 -0.000000 -0.000000 -0.000000 +1.000000

Fridays

Yes No  No  Yes Full $   No  No  Thai    30 - 60 No

No leaf node

Yes No  Yes Yes Full $   No  No  Thai    10 - 30 Yes

Yes leaf node

Yes Yes Yes Yes Full $   No  No  Burger  30 - 60 Yes

Yes leaf node
```