

# **EAI 320**

## Practical Assignment 5

Concept by Dr J. Dabrowski

Compiled, edited, and reviewed by Johan Langenhoven

08 March 2018

---

## Scenario

In this assignment, you are required to implement and test a decision tree building algorithm. In particular, the Iterative Dichotomiser 3 (ID3) decision tree building algorithm is considered<sup>1</sup>. For more information on the ID3 algorithm, consider the following sources see the Additional Material section at the end of this document. Instructions for the practical are as follows:

- Implement the Iterative Dichotomiser 3 (ID3) decision tree building algorithm in Python. As usual, you must implement a generic algorithm. That is, your function should be able to handle any dataset presented to it in a particular format.
- Test your ID3 algorithm on the restaurant dataset presented in the prescribed book. This dataset is presented in Chapter 18 and is reproduced as follows:

Example	Attributes										Target
	<i>Alt</i>	<i>Bar</i>	<i>Fri</i>	<i>Hun</i>	<i>Pat</i>	<i>Price</i>	<i>Rain</i>	<i>Res</i>	<i>Type</i>	<i>Est</i>	<i>WillWait</i>
$X_1$	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>French</i>	<i>0–10</i>	<i>T</i>
$X_2$	<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>30–60</i>	<i>F</i>
$X_3$	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>Some</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Burger</i>	<i>0–10</i>	<i>T</i>
$X_4$	<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>10–30</i>	<i>T</i>
$X_5$	<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>Full</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>French</i>	<i>&gt;60</i>	<i>F</i>
$X_6$	<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$</i>	<i>T</i>	<i>T</i>	<i>Italian</i>	<i>0–10</i>	<i>T</i>
$X_7$	<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>None</i>	<i>\$</i>	<i>T</i>	<i>F</i>	<i>Burger</i>	<i>0–10</i>	<i>F</i>
$X_8$	<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$</i>	<i>T</i>	<i>T</i>	<i>Thai</i>	<i>0–10</i>	<i>T</i>
$X_9$	<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>Full</i>	<i>\$</i>	<i>T</i>	<i>F</i>	<i>Burger</i>	<i>&gt;60</i>	<i>F</i>
$X_{10}$	<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>Italian</i>	<i>10–30</i>	<i>F</i>
$X_{11}$	<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>None</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>0–10</i>	<i>F</i>
$X_{12}$	<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Burger</i>	<i>30–60</i>	<i>T</i>

- Use the decision tree generated by the ID3 algorithm to make decisions on the following sample two examples. Show the decision path that is followed through the tree.

$$\begin{aligned}X_1 = \{ & Alt = No, Bar = No, Fri = Yes, \\ & Hun = Yes, Pat = Full, Price = \$ \$, \\ & Rain = Yes, Res = No, Type = Thai, Est = 10 - 30 \}\end{aligned}$$

$$\begin{aligned}X_2 = \{ & Alt = Yes, Bar = No, Fri = No, \\ & Hun = Yes, Pat = Full, Price = \$ \$, \\ & Rain = Yes, Res = No, Type = Italian, Est = 30 - 60 \}\end{aligned}$$

$$\begin{aligned}X_3 = \{ & Alt = Yes, Bar = No, Fri = Yes, \\ & Hun = Yes, Pat = Some, Price = \$, \\ & Rain = No, Res = No, Type = Burger, Est = 0 - 10 \}\end{aligned}$$

---

<sup>1</sup>The decision tree learning algorithm in the prescribed book is a variation of the ID3 algorithm

---

## Additional Material

- [1] T. Mitchell, Machine Learning, ser. McGraw-Hill international editions - computer science series. McGraw-Hill Education, 1997, ch. 3. [Online]. Available: <http://www.cs.princeton.edu/courses/archive/spr07/cos424/papers/mitchell-dectrees.pdf>
- [2] J. R. Quinlan, “Induction of decision trees”, Machine learning, vol. 1, no. 1, pp. 81-106, 1986. [Online]. Available: [https://wwwold.cs.umd.edu/class/fall2009/cmsc828r/PAPERS/fulltext\\_Quinlan\\_Ashwin\\_Kumar.pdf](https://wwwold.cs.umd.edu/class/fall2009/cmsc828r/PAPERS/fulltext_Quinlan_Ashwin_Kumar.pdf)
- [3] J. W. Grzymala-Busse, “Selected algorithms of machine learning from examples”, Fundam. Inform., vol. 18, pp. 193-207, 1993. [Online]. Available: <http://people.eecs.ku.edu/~jerzy/j24-sel.pdf>
- [4] Wikipedia, “Id3 algorithm”, [https://en.wikipedia.org/wiki/ID3\\_algorithm](https://en.wikipedia.org/wiki/ID3_algorithm).

If you can't find these papers by following the link, simply search the title of the paper on Google Scholar. There are also a couple of good video tutorials on Youtube.

---

# Report

You have to write a short technical report for this assignment. Your report must be written in  $\text{\LaTeX}$ . In the report you will give your results as well as provide a discussion on the results. Make sure to follow the guidelines as set out in the separate questions to form a complete report.

Reports will only be handed in as digital copies, but a hard copy plagiarism statement needs to be handed in at the following week's practical session (on the final day of the practical submission).

## Deliverable

- Write a technical report on your finding for this assignment.
- Include your code in the digital submission as an appendix.

## Instructions

- All reports must be in PDF format and be named `report.pdf`.
- Place the software in a folder called `SOFTWARE` and the report in a folder called `REPORT`.
- Add the folders to a zip-archive and name it `EAI320_prac5_studnr.zip`.
- All reports and simulation software must be e-mailed to **up.eai320@gmail.com** no later than 17:00 on 23 March 2018. No late submissions will be accepted.
- Use the following format for the subject header for the email: `EAI 320 Prac 5 - studnr`.
- Bring your plagiarism statements to the practical session on Thursday, 22 March 2018, where they will be collected.
- Submit your report online on ClickUP using the TurnItIn link.

## Additional Instructions

- Do not copy! The copier and the copyee (of software and/or documentation) will receive zero for both the software and the documentation.
- For any questions or appointments email me at **up.eai320@gmail.com**.
- Make sure that you discuss the results that are obtained. This is a large part of writing a technical report.

---

# Marking

Your report will be marked as follow:

- 60% will be awarded for the full implementation of the practical and the subsequent results in the report. For partially completed practicals, marks will be awarded as seen fit by the marker. **Commented code allows for easier marking!**
- 40% will be awarded for the overall report. This includes everything from the report structure, grammar and discussion of results. The discussion will be the bulk of the marks awarded.