



Project title

E-Book vs Book
Computing Intelligence subject

By

Aphinan Aphiwong
6231501076

Computer Engineering

School of Information Technology

MaeFahLuang university

Copyright © Mae Fah Luang university

Abstract

Technology has now played a role in our daily lives, so we have wondered how much of everyday items used by most people have used technology in daily life, so we have chosen to study physical books and e-books.

E-books are becoming more and more every day, so we'd like to know what kind of books most people prefer to read by surveying.

We studied by collecting data from online questionnaires and then analyzed the data to calculate the rate of chance.

Content

Topic	Page
Abstract	2
CHAPTER 1	4
1.1 TITLE	4
1.2 OBJECTIVE	4
1.3 SUMMARY OF ATTRIBUTES AND CLASS	4
1.4 Table OF (ALL) GATHERED DATA	5
CHAPTER 2	9
Calculation to find root node of decision tree :	9
Parameters	17
Accuracy	19
Precision	19
Recall	19
CHAPTER 3	20
Parameters	21
Accuracy	23
Precision	23
Recall	24
CHAPTER 4	24
CHAPTER 5	25
APPENDIX	26
REFERENCE (Questionnaire)	27

CHAPTER 1

INTRODUCTION

1.1 TITLE

-Book and Ebook

1.2 OBJECTIVE

-To know how many people prefer or read between books and e-book.

1.3 SUMMARY OF ATTRIBUTES AND CLASS

- 1) **Types of books** : comic,fiction,handbook,magazine
- 2) **Cover page** : beautiful, don't care
- 3) **Age** : less than 13, 13-19, more than 19
- 4) **Prices** : less than 300, 300-500, more than 500
- 5) **Do you think E-book is convenient than Physical books?** : yes, no

1.4 Table OF (ALL) GATHERED DATA

- All information that has been collected is 116 samples.

Row No.	Which do yo...	Which type ...	Choose fro...	Your age	Range of bo...	Do you think...
1	E-book	Fiction	Beautiful	More than 19	300-500 baht	Yes
2	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
3	books	Comic	Beautiful	More than 19	300-500 baht	Yes
4	E-book	Comic	Beautiful	More than 19	300-500 baht	Yes
5	books	Comic	Beautiful	More than 19	less than 300...	Yes
6	books	Comic	Don't care	More than 19	300-500 baht	No
7	E-book	Fiction	Don't care	More than 19	less than 300...	Yes
8	E-book	Comic	Beautiful	13-19	less than 300...	Yes
9	books	Fiction	Beautiful	More than 19	less than 300...	Yes
10	books	Comic	Don't care	More than 19	less than 300...	Yes
11	books	Comic	Don't care	More than 19	less than 300...	Yes
12	books	Comic	Don't care	More than 19	less than 300...	Yes
13	E-book	Comic	Beautiful	13-19	less than 300...	Yes
14	books	Comic	Don't care	More than 19	less than 300...	No
15	E-book	Comic	Beautiful	13-19	less than 300...	Yes
16	books	Comic	Beautiful	More than 19	300-500 baht	No
17	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
18	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
19	E-book	Fiction	Beautiful	More than 19	300-500 baht	Yes
20	books	Fiction	Beautiful	More than 19	less than 300...	No
21	books	Comic	Beautiful	More than 19	less than 300...	Yes
22	books	Handbook	Don't care	More than 19	300-500 baht	Yes
23	books	Fiction	Beautiful	More than 19	300-500 baht	No
24	E-book	Handbook	Don't care	More than 19	300-500 baht	Yes
25	E-book	Fiction	Beautiful	More than 19	300-500 baht	Yes
26	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
27	books	Fiction	Beautiful	More than 19	less than 300...	Yes
28	E-book	Handbook	Beautiful	13-19	less than 300...	Yes
29	E-book	Fiction	Beautiful	More than 19	less than 300...	Yes
30	books	Comic	Beautiful	More than 19	less than 300...	No
31	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
32	E-book	Fiction	Beautiful	13-19	less than 300...	Yes
33	books	Magazine	Beautiful	More than 19	less than 300...	Yes
34	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
35	E-book	Magazine	Beautiful	13-19	less than 300...	Yes
36	books	Comic	Beautiful	13-19	less than 300...	Yes

Row No.	Which do yo...	Which type ...	Choose fro...	Your age	Range of bo...	Do you think...
37	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
38	books	Comic	Beautiful	More than 19	less than 300...	Yes
39	books	Comic	Beautiful	More than 19	less than 300...	Yes
40	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
41	books	Comic	Beautiful	More than 19	less than 300...	No
42	books	Fiction	Beautiful	More than 19	less than 300...	No
43	E-book	Magazine	Beautiful	More than 19	less than 300...	Yes
44	books	Comic	Beautiful	More than 19	less than 300...	Yes
45	books	Comic	Beautiful	More than 19	less than 300...	Yes
46	books	Fiction	Beautiful	More than 19	300-500 baht	No
47	books	Fiction	Beautiful	13-19	less than 300...	Yes
48	books	Comic	Beautiful	13-19	less than 300...	Yes
49	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
50	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
51	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
52	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
Row No.	Which do yo...	Which type ...	Choose fro...	Your age	Range of bo...	Do you think...
55	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
56	books	Comic	Beautiful	13-19	less than 300...	Yes
57	books	Fiction	Beautiful	More than 19	less than 300...	Yes
58	books	Comic	Beautiful	13-19	300-500 baht	Yes
59	E-book	Fiction	Beautiful	More than 19	300-500 baht	Yes
60	E-book	Magazine	Beautiful	More than 19	less than 300...	Yes
61	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
62	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
63	books	Comic	Beautiful	More than 19	300-500 baht	Yes
64	books	Fiction	Beautiful	13-19	less than 300...	Yes
65	books	Handbook	Don't care	More than 19	more than 50...	Yes
66	E-book	Comic	Don't care	More than 19	300-500 baht	Yes
67	E-book	Fiction	Don't care	More than 19	300-500 baht	Yes
68	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
69	books	Comic	Beautiful	More than 19	less than 300...	Yes
70	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
71	books	Comic	Beautiful	less than 13	less than 300...	No
72	E-book	Fiction	Beautiful	More than 19	less than 300...	Yes

Row No.	Which do yo...	Which type ...	Choose fro...	Your age	Range of bo...	Do you think...
73	books	Magazine	Don't care	More than 19	300-500 baht	No
74	E-book	Fiction	Beautiful	More than 19	less than 300...	Yes
75	books	Comic	Beautiful	More than 19	300-500 baht	No
76	books	Comic	Beautiful	13-19	less than 300...	No
77	books	Magazine	Don't care	More than 19	less than 300...	Yes
78	E-book	Fiction	Don't care	More than 19	less than 300...	Yes
79	books	Comic	Beautiful	More than 19	less than 300...	Yes
80	books	Magazine	Beautiful	More than 19	300-500 baht	No
81	E-book	Comic	Beautiful	13-19	less than 300...	Yes
82	E-book	Fiction	Don't care	13-19	less than 300...	Yes
83	books	Comic	Beautiful	More than 19	more than 50...	Yes
84	books	Comic	Beautiful	More than 19	less than 300...	Yes
85	books	Comic	Don't care	less than 13	300-500 baht	Yes
86	books	Fiction	Don't care	less than 13	less than 300...	No
87	E-book	Magazine	Beautiful	less than 13	300-500 baht	Yes
88	books	Comic	Beautiful	less than 13	more than 50...	Yes
89	books	Magazine	Beautiful	less than 13	300-500 baht	Yes
Row No.	Which do yo...	Which type ...	Choose fro...	Your age	Range of bo...	Do you think...
91	books	Magazine	Don't care	More than 19	300-500 baht	Yes
92	E-book	Fiction	Beautiful	More than 19	300-500 baht	Yes
93	E-book	Fiction	Don't care	More than 19	300-500 baht	Yes
94	books	Fiction	Beautiful	More than 19	less than 300...	No
95	books	Magazine	Beautiful	More than 19	less than 300...	Yes
96	books	Comic	Beautiful	More than 19	less than 300...	Yes
97	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
98	E-book	Handbook	Beautiful	More than 19	300-500 baht	Yes
99	books	Fiction	Don't care	More than 19	less than 300...	No
100	E-book	Magazine	Don't care	More than 19	300-500 baht	Yes
101	books	Comic	Beautiful	More than 19	less than 300...	Yes
102	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
103	books	Fiction	Beautiful	13-19	300-500 baht	Yes
104	books	Fiction	Don't care	More than 19	less than 300...	Yes
105	books	Fiction	Beautiful	More than 19	less than 300...	Yes
106	E-book	Fiction	Don't care	More than 19	more than 50...	Yes
107	E-book	Fiction	Don't care	More than 19	300-500 baht	Yes
108	E-book	Comic	Don't care	More than 19	less than 300...	Yes

109	books	Fiction	Beautiful	More than 19	300-500 baht	Yes
110	E-book	Handbook	Beautiful	More than 19	300-500 baht	Yes
111	E-book	Comic	Don't care	More than 19	less than 300...	Yes
112	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
113	books	Comic	Don't care	More than 19	less than 300...	No
114	E-book	Comic	Beautiful	More than 19	less than 300...	Yes
115	books	Fiction	Don't care	less than 13	less than 300...	Yes
116	E-book	Magazine	Beautiful	More than 19	more than 50...	Yes

CHAPTER 2

CALCULATION FOR DECISION TREE CLASSIFICATION

Calculation to find root node of decision tree :

$S = 116$	$I(S_1, S_2) = I(68, 48)$
$\text{DATA} \Rightarrow S_1 = 68 \text{ (book)}$	$= \left(\frac{-68}{116} \log_2 \frac{68}{116} \right) - \left(\frac{48}{116} \log_2 \frac{48}{116} \right)$
$\Rightarrow S_2 = 48 \text{ (e-book)}$	$\approx 0.938449 \rightarrow 0.93845$

There are 116 samples
Choose book 68
Choose ebook 48

Step1: Find entropy of each attribute

TYPE OF BOOK :	
(S1) CASE: COMIC B(33) E(21)	$I(33, 21) = \left(\frac{-33}{54} \log_2 \frac{33}{54} \right) - \left(\frac{21}{54} \log_2 \frac{21}{54} \right) = 0.9640993 \rightarrow 0.96408$
(C2) CASE: FICTION B(26) E(16)	$I(26, 16) = \left(\frac{-26}{42} \log_2 \frac{26}{42} \right) - \left(\frac{16}{42} \log_2 \frac{16}{42} \right) \approx 0.9583118 \rightarrow 0.95831$
(B) CASE: HANDBOOK B(2) E(4)	$I(2, 4) = \left(\frac{-2}{6} \log_2 \frac{2}{6} \right) - \left(\frac{4}{6} \log_2 \frac{4}{6} \right) = 0.918296 \rightarrow 0.91830$
(M4) CASE: MAGAZINE B(9) E(7)	$I(9, 7) = \left(\frac{-9}{14} \log_2 \frac{9}{14} \right) - \left(\frac{7}{14} \log_2 \frac{7}{14} \right) = 1 \rightarrow 1.00000$

COVER :	
(B1) CASE: BEAUTY B(52) E(36)	$I(52, 36) = \left(\frac{-52}{88} \log_2 \frac{52}{88} \right) - \left(\frac{36}{88} \log_2 \frac{36}{88} \right) = 0.9960206 \rightarrow 0.99602$
(C8) CASE: DON'T CARE B(16) E(12)	$I(16, 12) = \left(\frac{-16}{28} \log_2 \frac{16}{28} \right) - \left(\frac{12}{28} \log_2 \frac{12}{28} \right) = 0.986228 \rightarrow 0.98623$

AGE :	
(L) CASE: LESS 15 B(6) E(1)	$I(6, 1) = \left(\frac{-6}{7} \log_2 \frac{6}{7} \right) - \left(\frac{1}{7} \log_2 \frac{1}{7} \right) = 0.5416927 \rightarrow 0.54167$
(15-19) CASE: 15-19 B(9) E(4)	$I(9, 4) = \left(\frac{-9}{13} \log_2 \frac{9}{13} \right) - \left(\frac{4}{13} \log_2 \frac{4}{13} \right) = 0.4995025464 \rightarrow 0.4995025464$
(M19) CASE: MORE 19 B(54) E(38)	$I(54, 38) = \left(\frac{-54}{92} \log_2 \frac{54}{92} \right) - \left(\frac{38}{92} \log_2 \frac{38}{92} \right) = 0.9780109 \rightarrow 0.9780109$

PRICE :	
(<10) CASE: <300 B(19) E(11)	$I(19, 11) = \left(\frac{-19}{30} \log_2 \frac{19}{30} \right) - \left(\frac{11}{30} \log_2 \frac{11}{30} \right) = 0.9905537 \rightarrow 0.99056$
(100-500) CASE: 300-500 B(26) E(15)	$I(26, 15) = \left(\frac{-26}{41} \log_2 \frac{26}{41} \right) - \left(\frac{15}{41} \log_2 \frac{15}{41} \right) = 0.9434351 \rightarrow 0.94344$
(>500) CASE: >500 B(3) E(2)	$I(3, 2) = \left(\frac{-3}{5} \log_2 \frac{3}{5} \right) - \left(\frac{2}{5} \log_2 \frac{2}{5} \right) = 0.97095029 \rightarrow 0.97095$

CONVENIENT :

$$(d) \text{ CASE: YES } BC(50) \quad E(48) \quad I(50,48) = \left(-\frac{50}{48} \log_2 \frac{50}{48} \right) - \left(\frac{48}{48} \log_2 \frac{48}{48} \right) = 0.999945 \xrightarrow{\text{round}} 0.99990$$

$$(e) \text{ CASE: NO } BL(8) \quad E(0) \quad I(18,0) = \left(-\frac{18}{18} \log_2 \frac{18}{18} \right) - \left(\frac{0}{18} \log_2 \frac{0}{18} \right) = 0 \xrightarrow{\text{round}} 0.00000$$

$$E(\text{TYPE OF BOOK}) = \left[\frac{54(I_{CS,10}) + 42(I_{CS,15}) + 6(I_{CS,45}) + 14(I_{CS,35})}{116} \right] = \left[54(0.96409) + 42(0.95871) + 6(0.91830) + 14(1.00000) \right] \\ = 0.964103 \xrightarrow{\text{round}} 0.96410$$

$$E(\text{COVER}) = \left[\frac{88(I_{CS,36}) + 28(I_{OB,35})}{116} \right] = \left[\frac{88(0.93602) + 28(0.98523)}{116} \right] = 0.978243 \xrightarrow{\text{round}} 0.97824$$

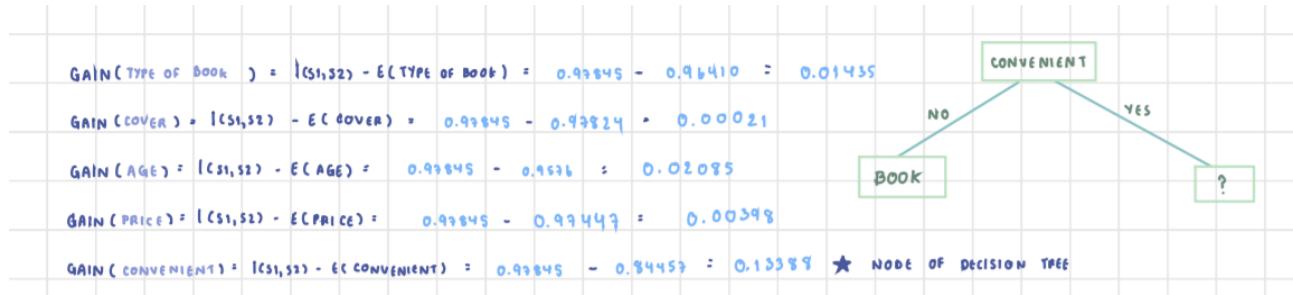
$$E(\text{AGE}) = \left[\frac{9(I_{16,17}) + 17(I_{19,45}) + 92(I_{30+,35})}{116} \right] = \left[\frac{9(0.54167) + 17(0.99750) + 92(0.99803)}{116} \right] = 0.9596 \xrightarrow{\text{round}} 0.9596$$

$$E(\text{AGE}) = \left[\frac{9(I_{16,17}) + 17(I_{19,45}) + 92(I_{30+,35})}{116} \right] = \left[\frac{9(0.54167) + 17(0.99750) + 92(0.99803)}{116} \right] = 0.9596 \xrightarrow{\text{round}} 0.9596$$

$$E(\text{PRICE}) = \left[\frac{30(I_{CS,35}) + 41(I_{CS,15}) + 5(I_{10+,15})}{116} \right] = \left[\frac{30(0.99056) + 41(0.94444) + 5(0.97095)}{116} \right] = 0.974474 \xrightarrow{\text{round}} 0.97447$$

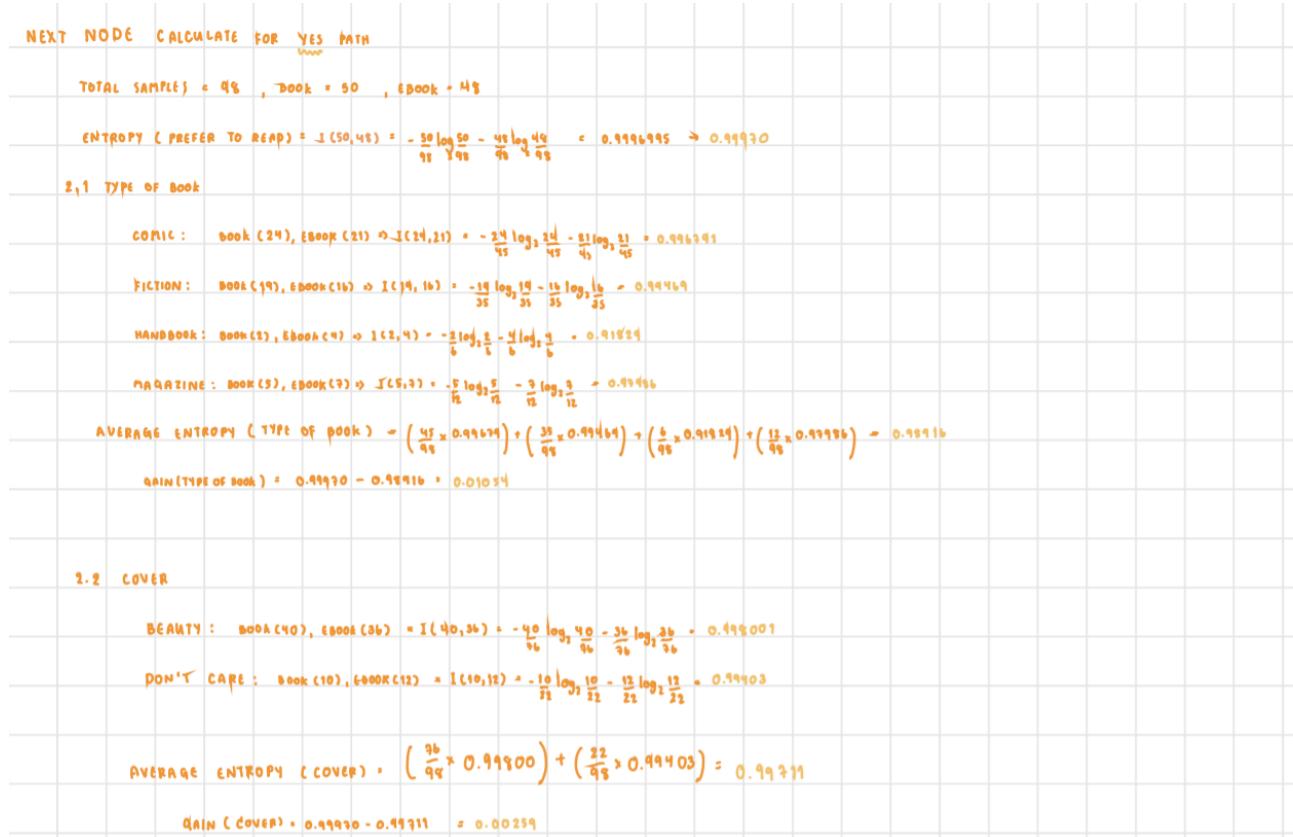
$$E(\text{CONVENIENT}) = \left[\frac{98(I_{50,48}) + 18(I_{18,0})}{116} \right] = \left[\frac{98(0.99990) + 18(0.00000)}{116} \right] = 0.844574 \xrightarrow{\text{round}} 0.84457$$

Step 2: Find gain of each attribute



So, **convenient** is suitable to be root node of decision tree

Step 3: Find next node for path YES



2.3 AGE

$$\text{LESS THAN } 300 : \text{BOOK}(4), \text{EBOOK}(1) \Rightarrow I(4,1) = -\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5} = 0.921928 \rightarrow 0.92193$$

$$300-500 : \text{BOOK}(9), \text{EBOOK}(9) \Rightarrow I(9,9) = -\frac{9}{16} \log_2 \frac{9}{16} - \frac{9}{16} \log_2 \frac{9}{16} = 0.988699 \rightarrow 0.98860$$

$$\text{MORE THAN } 500 : \text{BOOK}(39), \text{EBOOK}(38) \Rightarrow I(39,38) = -\frac{39}{77} \log_2 \frac{39}{77} - \frac{38}{77} \log_2 \frac{38}{77} = 0.994828 \rightarrow 0.99488$$

$$\text{AVERAGE ENTROPY(AGE)} = \left(\frac{5}{98} \times 0.92193\right) + \left(\frac{16}{98} \times 0.98860\right) + \left(\frac{77}{98} \times 0.99488\right) = 0.98396$$

$$\text{GAIN(AGE)} = 0.44970 - 0.98396 = 0.01584$$

2.4 PRICE

$$\text{LESS THAN } 300 : \text{BOOK}(23), \text{EBOOK}(31) \Rightarrow I(23,31) = -\frac{23}{54} \log_2 \frac{23}{54} - \frac{31}{54} \log_2 \frac{31}{54} = 0.99813$$

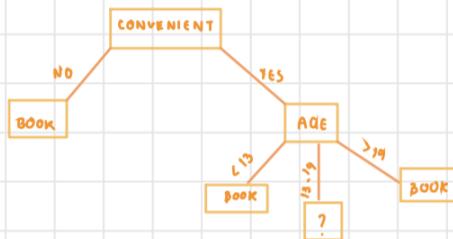
$$300-500 : \text{BOOK}(19), \text{EBOOK}(15) \Rightarrow I(19,15) = -\frac{19}{34} \log_2 \frac{19}{34} - \frac{15}{34} \log_2 \frac{15}{34} = 0.98999$$

$$\text{MORE THAN } 500 : \text{BOOK}(2), \text{EBOOK}(2) \Rightarrow I(2,2) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = 0.97095$$

$$\text{AVERAGE ENTROPY(PRICE)} = \left(\frac{5}{98} \times 0.99813\right) + \left(\frac{34}{98} \times 0.98999\right) + \left(\frac{2}{98} \times 0.97095\right) = 0.99341$$

$$\text{GAIN(PRICE)} = 0.99930 - 0.99341 = 0.00579$$

THEREFOR : AGE IS NEXT NODE FOR YES PATH



So, age is the next node of this path

Step 4: Find next node for path 13-19

NEXT NODE CALCULATE FOR 13-19 PATH

TOTAL SAMPLE = 16 BOOK(7), EBOOK(9)

$$\text{ENTROPY} (13-19) = I(3,4) = -\frac{7}{16} \log_2 \frac{7}{16} - \frac{9}{16} \log_2 \frac{9}{16} = 0.98869$$

3.1 TYPE OF BOOK

$$\text{COMIC: } \text{BOOK}(4), \text{EBOOK}(4) \Rightarrow I(4,4) = -\frac{4}{8} \log_2 \frac{4}{8} - \frac{4}{8} \log_2 \frac{4}{8} = 1$$

$$\text{FICTION: } \text{BOOK}(3), \text{EBOOK}(2) \Rightarrow I(3,2) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.93095$$

$$\text{HANDBOOK: } \text{BOOK}(0), \text{EBOOK}(1) \Rightarrow I(1,0) = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

$$\text{MAGAZINE: } \text{BOOK}(0), \text{EBOOK}(2) \Rightarrow I(2,0) = -\frac{1}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2} = 0$$

$$\text{AVERAGE ENTROPY(TYPE)} = \left(\frac{8}{16} \times 1\right) + \left(\frac{5}{16} \times 0.93095\right) + \left(\frac{1}{16} \times 0\right) + \left(\frac{2}{16} \times 0\right) = 0.80342$$

$$\text{GAIN (TYPE OF BOOK)} = 0.98869 - 0.80342 = 0.18528$$

3.2 COVER

$$\text{BEAUTY: } \text{BOOK}(7), \text{EBOOK}(8) \Rightarrow I(3,8) = -\frac{7}{15} \log_2 \frac{7}{15} - \frac{8}{15} \log_2 \frac{8}{15} = 0.99679$$

$$\text{DON'T CARE: } \text{BOOK}(0), \text{EBOOK}(1) \Rightarrow I(0,1) = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$\text{AVERAGE ENTROPY(COVER)} = \left(\frac{15}{16} \times 0.99679\right) + \left(\frac{1}{16} \times 0\right) = 0.93449$$

$$\text{GAIN (COVER)} = 0.98869 - 0.93449 = 0.05421$$

3.3 PRICE

$$\text{LESS THAN 300: } \text{BOOK}(5), \text{EBOOK}(4) \Rightarrow I(5,4) = -\frac{5}{14} \log_2 \frac{5}{14} - \frac{4}{14} \log_2 \frac{4}{14} = 0.94023$$

$$\text{300-500: } \text{BOOK}(2), \text{EBOOK}(0) \Rightarrow I(2,0) = -\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2} = 0$$

$$\text{MORE THAN 500: } \text{BOOK}(0), \text{EBOOK}(0) \Rightarrow I(0,0) = -\frac{0}{0} \log_2 \frac{0}{0} - \frac{0}{0} \log_2 \frac{0}{0} = 0$$

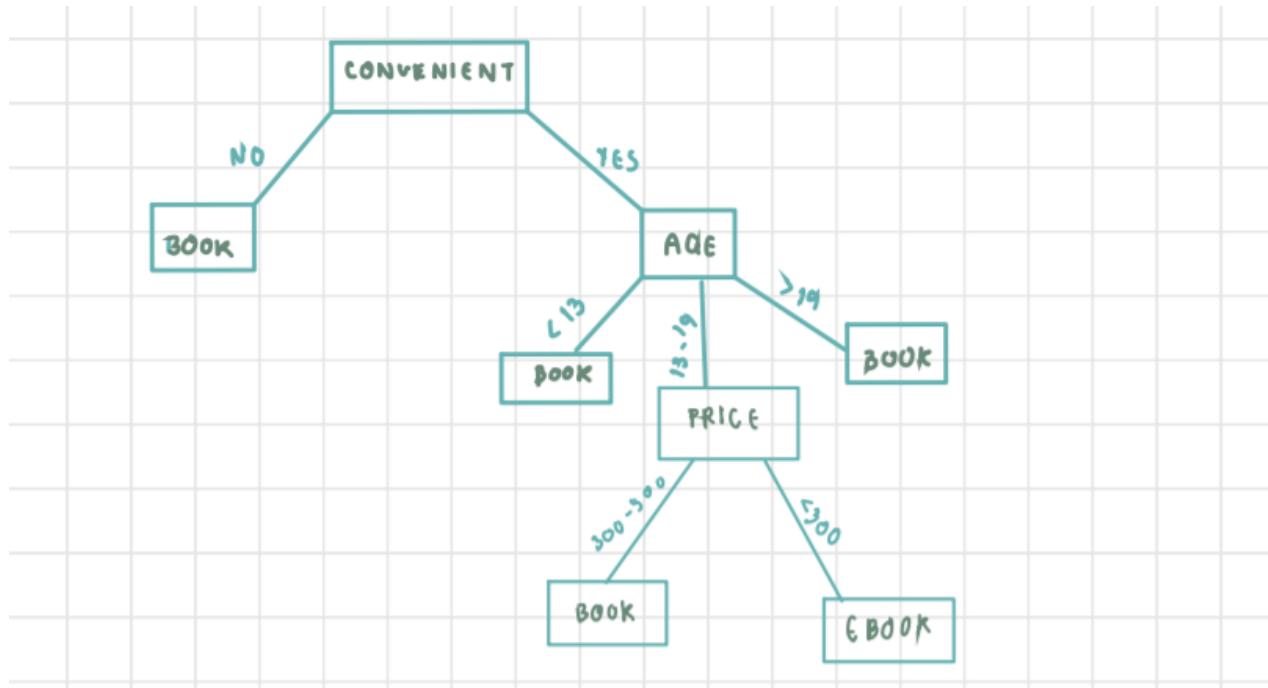
$$\text{AVERAGE ENTROPY (PRICE)} = \left(\frac{14}{16} \times 0.94023\right) + \left(\frac{2}{16} \times 0\right) + \left(\frac{0}{16} \times 0\right) = 0.82234$$

$$\text{GAIN (PRICE)} = 0.98869 - 0.82234 = 0.16596$$

∴ THEN PRICE IS NEXT NODE FOR 13-19 PATH

So, price is the next node of this path

The complete tree is shown as follows:



IF E-BOOK IS NOT CONVENIENT THAN PHYSICAL BOOK

THEN READ PHYSICAL BOOK

IF E-BOOK IS CONVENIENT THAN PHYSICAL BOOK

THEN IF YOUR AGE IS 13-19

THEN RANGE OF PRICE IS 300-500 BAHT

THEN READ PHYSICAL BOOK

THEN RANGE OF PRICE IS < 300 BAHT

THEN READ PHYSICAL BOOK

THEN IF YOUR AGE IS >19

THEN READ PHYSICAL BOOK

THEN IF YOUR AGE IS < 13

THEN READ PHYSICAL BOOK

First, import the data for the project in the workspace of RapidMiner Studio and use operator Nominal Cross Validation for training and testing data.

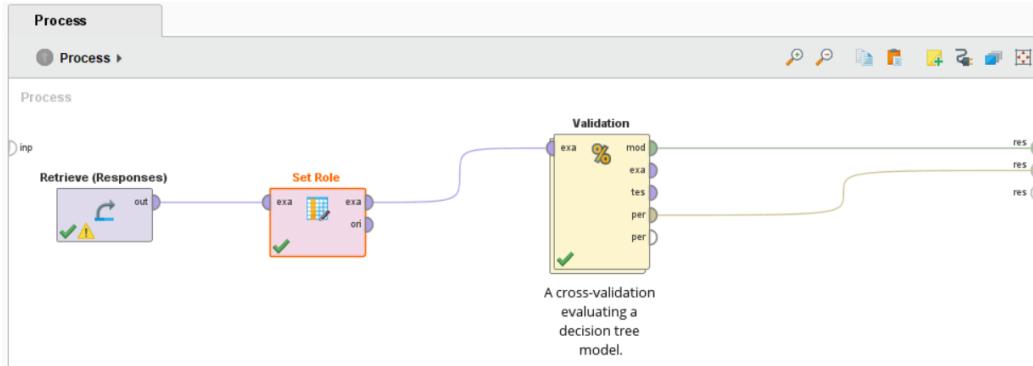


Figure 1: Operator in decision tree classification

In Validation, there are training and testing methods. And build the model of dataset (90% of the data by 5 times) as followed:

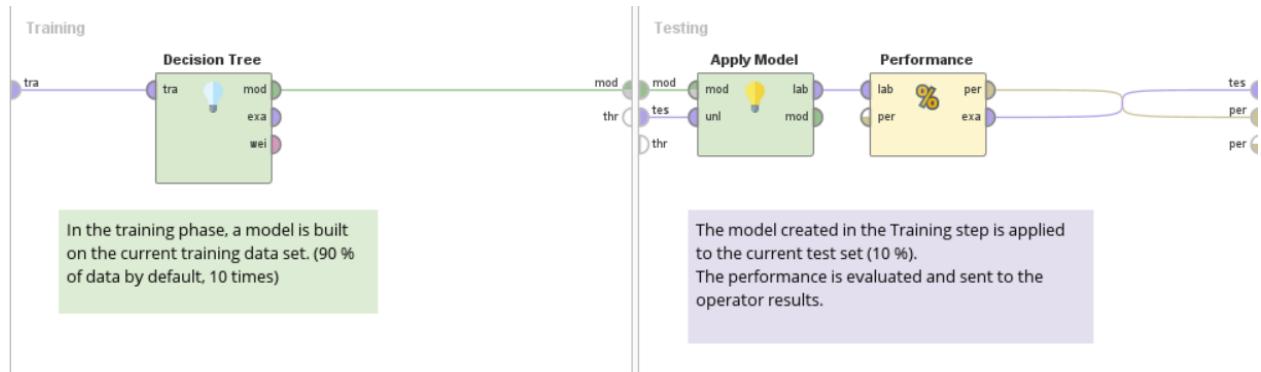


Figure 2: Inside the Nominal Cross Validation (Training and Testing)

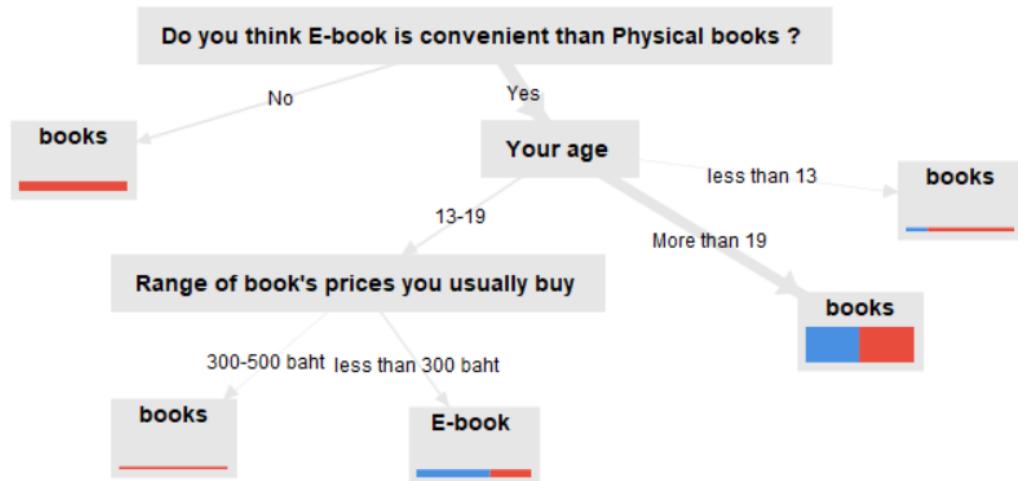


Figure 3: Decision Tree

Parameters

Set Role <p>attribute name: Which do you prefer to read ?</p> <p>target role: label</p> <p>set additional roles: Edit List (0)...</p>	Validation (Cross Validation) <p><input type="checkbox"/> split on batch attribute</p> <p><input type="checkbox"/> leave one out</p> <p>number of folds: 5</p> <p>sampling type: automatic</p> <p><input type="checkbox"/> use local random seed</p> <p><input checked="" type="checkbox"/> enable parallel execution</p>
Decision Tree <p>criterion: gain_ratio</p> <p>maximal depth: 10</p> <p><input checked="" type="checkbox"/> apply pruning</p> <p>confidence: 0.1</p> <p><input checked="" type="checkbox"/> apply prepruning</p> <p>minimal gain: 0.01</p> <p>minimal leaf size: 2</p> <p>minimal size for split: 4</p> <p>number of prepruning alternatives: 3</p>	

From the calculation, Performance Vector will describe for this calculation are as followed:

PerformanceVector

```

PerformanceVector:
accuracy: 50.15% +/- 7.83% (micro average: 50.00%)
ConfusionMatrix:
True: E-book books
E-book : 22 32
books: 26 36
precision: 62.02% +/- 15.34% (micro average: 58.06%) (positive class: books)
ConfusionMatrix:
True: E-book books
E-book : 22 32
books: 26 36
recall: 54.46% +/- 26.23% (micro average: 52.94%) (positive class: books)
ConfusionMatrix:
True: E-book books
E-book : 22 32
books: 26 36
AUC (optimistic): 0.780 +/- 0.136 (micro average: 0.780) (positive class: books)
AUC: 0.604 +/- 0.109 (micro average: 0.604) (positive class: books)
AUC (pessimistic): 0.448 +/- 0.167 (micro average: 0.448) (positive class: books)

```

Accuracy

accuracy: 52.64% +/- 6.17% (micro average: 52.59%)

	true E-book	true books	class precision
pred. E-book	18	25	41.86%
pred. books	30	43	58.90%
class recall	37.50%	63.24%	

Precision

precision: 58.94% +/- 3.21% (micro average: 58.90%) (positive class: books)

	true E-book	true books	class precision
pred. E-book	18	25	41.86%
pred. books	30	43	58.90%
class recall	37.50%	63.24%	

Recall

recall: 63.52% +/- 24.35% (micro average: 63.24%) (positive class: books)

	true E-book	true books	class precision
pred. E-book	18	25	41.86%
pred. books	30	43	58.90%
class recall	37.50%	63.24%	

CHAPTER 3

CALCULATION FOR NEURAL NETWORK CLASSIFICATION

First, import the data for the project in the workspace of RapidMiner Studio and use Set Role, Nominal to Numerical operator and use Nominal Cross Validation operator for training and testing data.

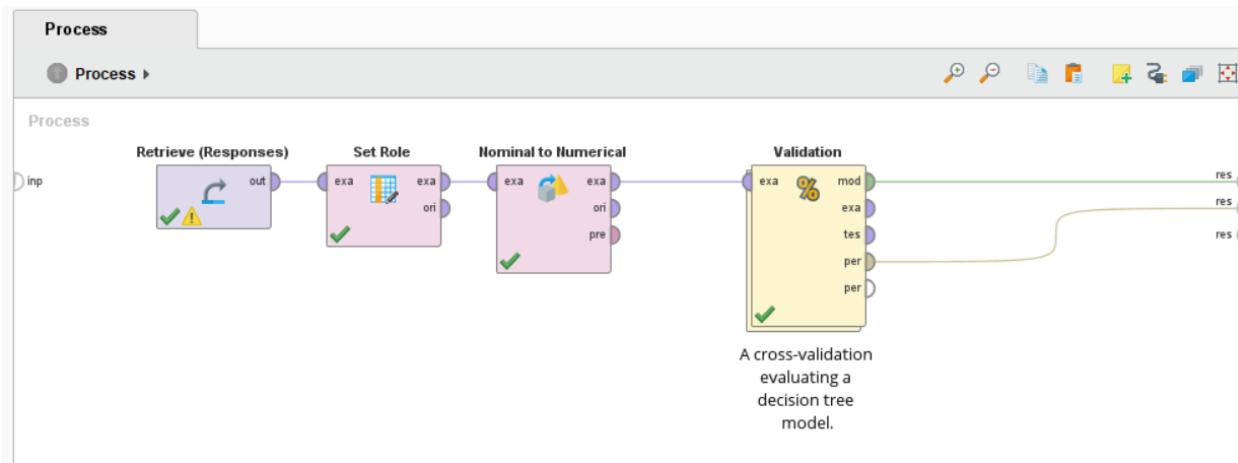


Figure 4: Operator in neuron network classification

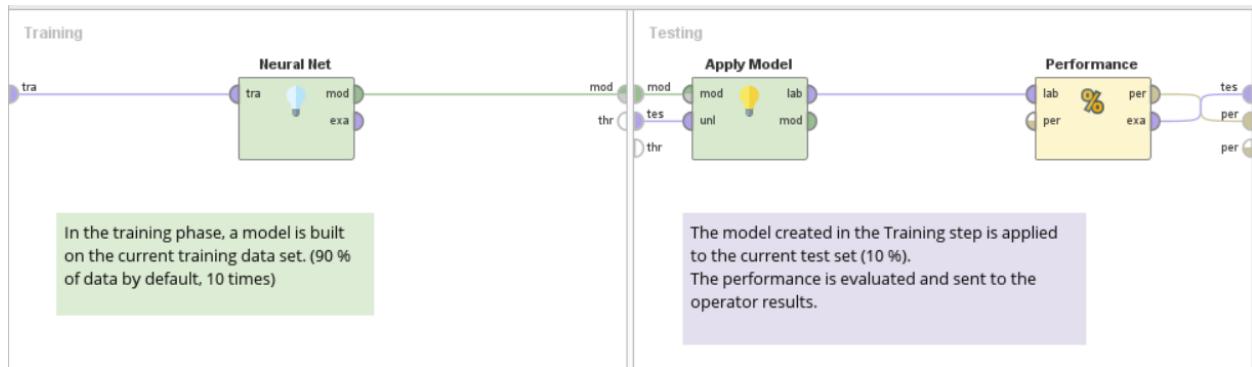


Figure 5: Inside the Nominal Cross Validation (Training and Testing)

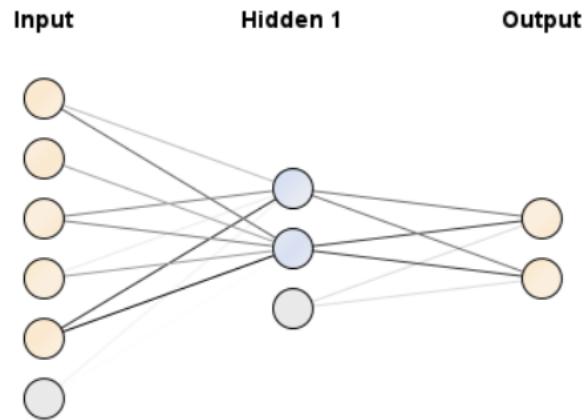
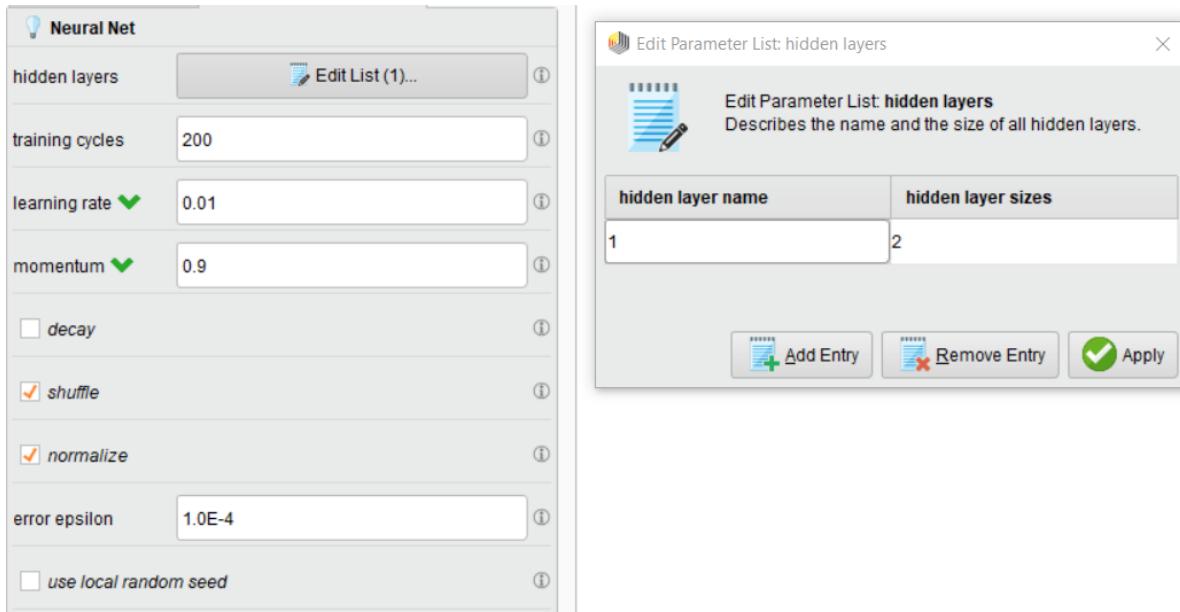


Figure 6: Neural Net

Parameters

<p>Set Role</p> <p>attribute name <input type="text" value="Which do you prefer to read ?"/></p> <p>target role <input type="text" value="label"/></p> <p>set additional roles <input style="width: 150px; height: 20px;" type="button" value="Edit List (0)..."/></p>	<p>Nominal to Numerical</p> <p><input type="checkbox"/> create view</p> <p>attribute filter type <input type="text" value="all"/></p> <p><input type="checkbox"/> invert selection</p> <p><input type="checkbox"/> include special attributes</p> <p>coding type <input type="text" value="unique integers"/></p>
<p>Validation (Cross Validation)</p> <p><input type="checkbox"/> split on batch attribute</p> <p><input type="checkbox"/> leave one out</p> <p>number of folds <input type="text" value="5"/></p> <p>sampling type <input type="text" value="automatic"/></p> <p><input type="checkbox"/> use local random seed</p> <p><input checked="" type="checkbox"/> enable parallel execution</p>	



In Validation, there are training and testing methods like Decision Tree classification. But in the Neural Network, we need to remove the Decision Tree and use Neural Net instead in Training. And build the model of the dataset (90% of the data by 5 times). The Performance Vector will describe for this calculation are as followed:

PerformanceVector

```
PerformanceVector:
accuracy: 66.45% +/- 10.45% (micro average: 66.38%)
ConfusionMatrix:
True: E-book books
E-book :      35      26
books : 13      42
precision: 77.40% +/- 13.84% (micro average: 76.36%) (positive class: books )
ConfusionMatrix:
True: E-book books
E-book :      35      26
books : 13      42
recall: 61.76% +/- 7.72% (micro average: 61.76%) (positive class: books )
ConfusionMatrix:
True: E-book books
E-book :      35      26
books : 13      42
AUC (optimistic): 0.718 +/- 0.097 (micro average: 0.718) (positive class: books )
AUC: 0.688 +/- 0.107 (micro average: 0.688) (positive class: books )
AUC (pessimistic): 0.658 +/- 0.117 (micro average: 0.658) (positive class: books )
```

Accuracy

accuracy: 66.45% +/- 10.45% (micro average: 66.38%)

	true E-book	true books	class precision
pred. E-book	35	26	57.38%
pred. books	13	42	76.36%
class recall	72.92%	61.76%	

Precision

precision: 77.40% +/- 13.84% (micro average: 76.36%) (positive class: books)

	true E-book	true books	class precision
pred. E-book	35	26	57.38%
pred. books	13	42	76.36%
class recall	72.92%	61.76%	

Recall

recall: 61.76% +/- 7.72% (micro average: 61.76%) (positive class: books)

	true E-book	true books	class precision
pred. E-book	35	26	57.38%
pred. books	13	42	76.36%
class recall	72.92%	61.76%	

CHAPTER 4

RESULT COMPARISON AND DISCUSSION

From the survey data in the title of Ebook vs Book with 116 samples, we can compare the result of Decision tree and Neural Network.

There is accuracy : 52.64% +/- 6.17% (micro average: 52.59%) in Decision Tree

There is accuracy : 66.45% +/- 10.45% (micro average 66.38%) in Neural Network

Decision tree

accuracy: 52.64% +/- 6.17% (micro average: 52.59%)

	true E-book	true books	class precision
pred. E-book	18	25	41.86%
pred. books	30	43	58.90%
class recall	37.50%	63.24%	

Neural Network

accuracy: 66.45% +/- 10.45% (micro average: 66.38%)

	true E-book	true books	class precision
pred. E-book	35	26	57.38%
pred. books	13	42	76.36%
class recall	72.92%	61.76%	

So, the Neural Network has accuracy more than Decision Tree for 13.81% because these methods can calculate the data like a human's brain can do.

Calculation of Neural Networks will be complicated for calculating.

CHAPTER 5

CONCLUSION AND SUGGESTION

Both analysis models can be used to analyze data.

But accuracy is not the same. The results showed that calculate for Decision tree classification had an accuracy of 52.64% and, calculate for Neural network classification had an accuracy of 66.45% from all 116 samples.

This can be seen that the calculation of the data with a Neural network is more accurate than the Decision tree.

The suggestion for this project is data collection. The above information is distributed similarly to the form of statistics.

Therefore, data collection should have a clear target. This includes increasing the number of samples.

However, since this project already has statistical data, it can be used as a database to be used for future data analysis.

APPENDIX

- 1) Types of books :** comic,fiction,handbook,magazine
- 2) Cover page :** beautiful, don't care
- 3) Age :** less than 13, 13-19, more than 19
- 4) Prices :** less than 300, 300-500, more than 500
- 5) Do you think E-book is convenient than Physical books? :** yes, no

REFERENCE (Questionnaire)

<https://docs.google.com/forms/d/1qggYqJHdpX6qBoqxEb4Jxh1QQBPEqNFZIIQ6o3bJrxc/edit>



What kinds of books (Book or E-Book)
do people choose to read more these
days

The form What kinds of books (Book or E-Book) do people choose to read more these days is no longer accepting responses.

Try contacting the owner of the form if you think this is a mistake.

Which type of books do you prefer ? (คุณชอบหนังสือประเภทไหน)

*

- Comic (หนังสือการ์ตูน)
- Fiction (หนังสือนิยาย)
- Handbook (หนังสือเรียน)
- Magazine (นิตยสาร)

Choose from cover page (เลือกจากหน้าปก) *

- Beautiful (หน้าปกสวย)
- Don't care (ไม่สนใจ)

Your age (อายุ) *

- less than 13 (น้อยกว่า 13 ปี)
- 13-19
- More than 19 (มากกว่า 19 ปี)

Range of book's prices you usually buy (ช่วงของราคาที่คุณมักจะซื้อ) *

- less than 300 baht (น้อยกว่า 300 บาท)
- 300-500 baht (300 - 500 บาท)
- more than 500 baht (มากกว่า 500 บาท)

Do you think E-book is convenient than Physical books ? (คุณคิดว่าหนังสืออิเล็กทรอนิกส์สะดวกกว่า หนังสือเล่มหรือไม่)

- Yes (ใช่)
- No (ไม่)

Which do you prefer to read ? (คุณชอบอ่านแบบไหน) *

- E-book (หนังสืออิเล็กทรอนิกส์)
- books (หนังสือเล่ม)