



RMUTT

www.rmutt.ac.th ราชมนเทียรชัยบุรี



Prototyped Mobile Application for Rice Product Prediction using Classification Techniques

Asst.Pro.Dr.Uraiwan Inyaem
Department of Computer Science
Faculty of Science and Technology

RMUTT

Rajamangala University of Technology Thanyaburi

Agenda



1. Research Background
2. Related Works
3. Methodologies
4. Experimental Result
5. Application
6. Application Assessment
7. Q & A



1. Research Background





2. Related Works

Related Works

1. Data Mining
2. Classification
3. Decision Tree
4. Naïve Bayes

Data Mining



- Introduction to Data Mining
- Stages of Data Mining Process
 - a. Data Cleaning
 - b. Data Integration
 - c. Data Selection
 - d. Data Transformation
 - e. Data Mining
 - f. Pattern Evaluation
 - g. Knowledge Representation

Data Mining



Data Mining

06-Jun-19

Data mining techniques

Classification

Clustering

Regression

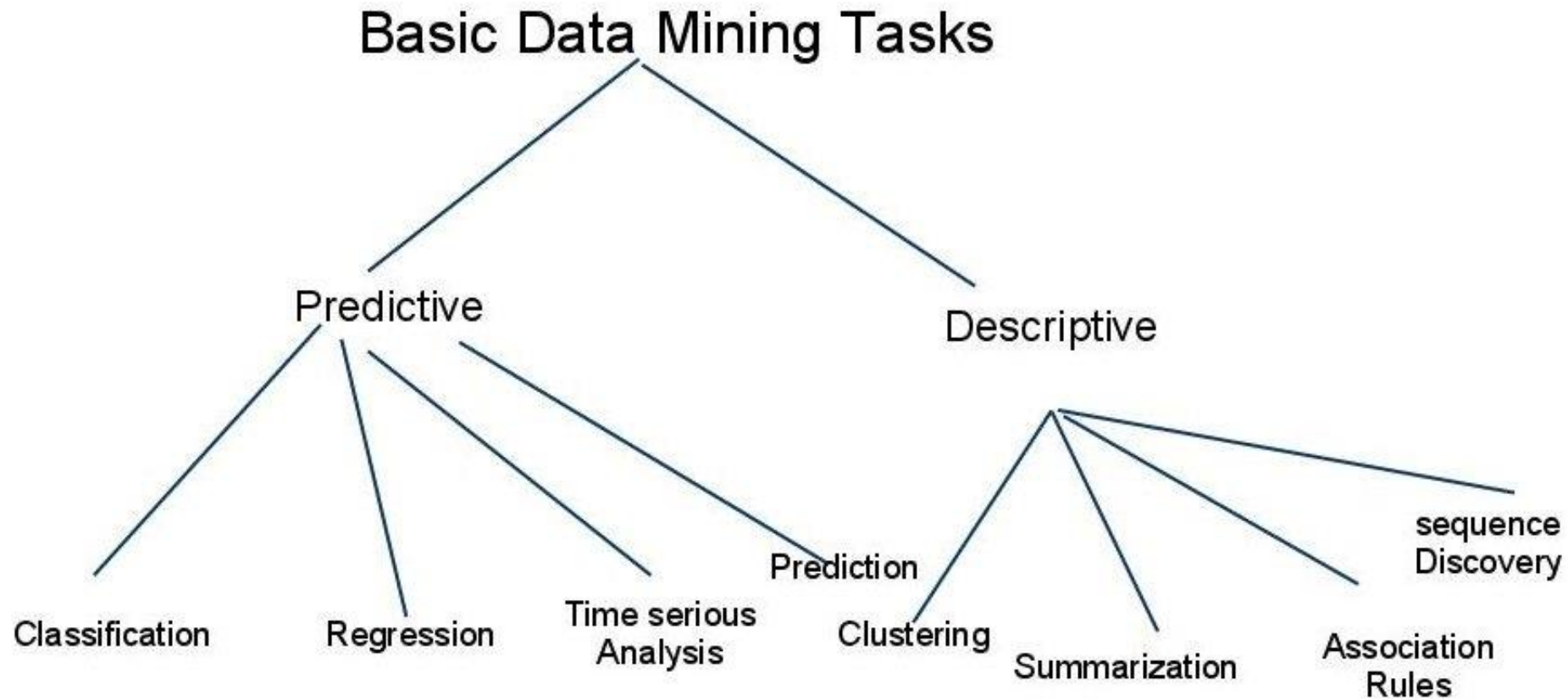
Outer

Sequential
Patterns

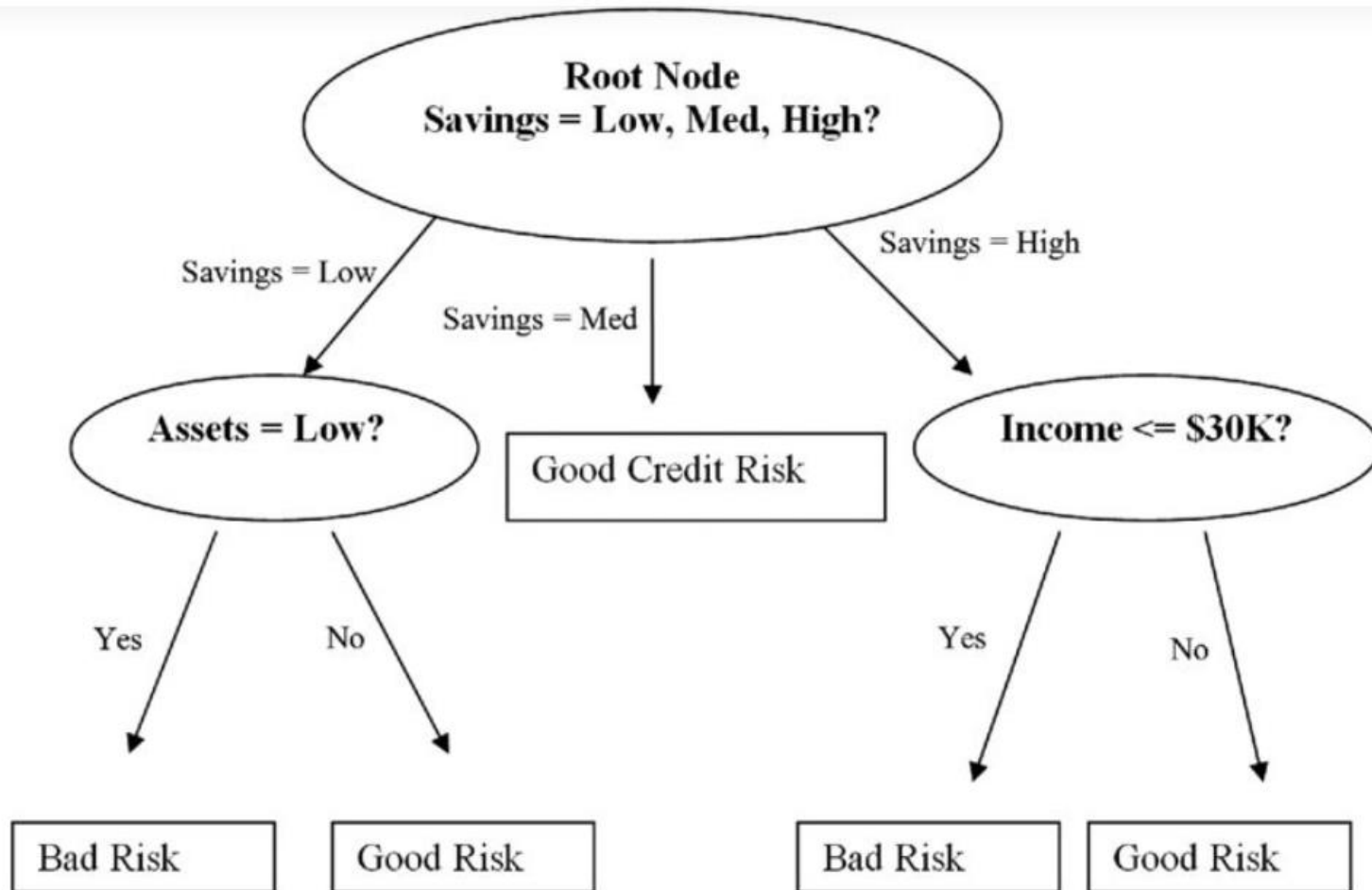
Prediction

Association
Rules

Data Mining



Decision Tree



Common terms used with Decision trees



- 1. Root Node:** It represents entire population or sample and this further gets divided into two or more homogeneous sets.
- 2. Splitting:** It is a process of dividing a node into two or more sub-nodes.
- 3. Decision Node:** When a sub-node splits into further sub-nodes, then it is called decision node.
- 4. Leaf/ Terminal Node:** Nodes do not split is called Leaf or Terminal node.
- 5. Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. You can say opposite process of splitting.
- 6. Branch / Sub-Tree:** A sub section of entire tree is called branch or sub-tree.
- 7. Parent and Child Node:** A node, which is divided into sub-nodes is called parent node of sub-nodes whereas sub-nodes are the child of parent node.

Decision Tree

$$H(S) = \sum_{c \in C} -p(c) \log_2 p(c)$$

$$IG(A, S) = H(S) - \sum_{t \in T} p(t)H(t)$$

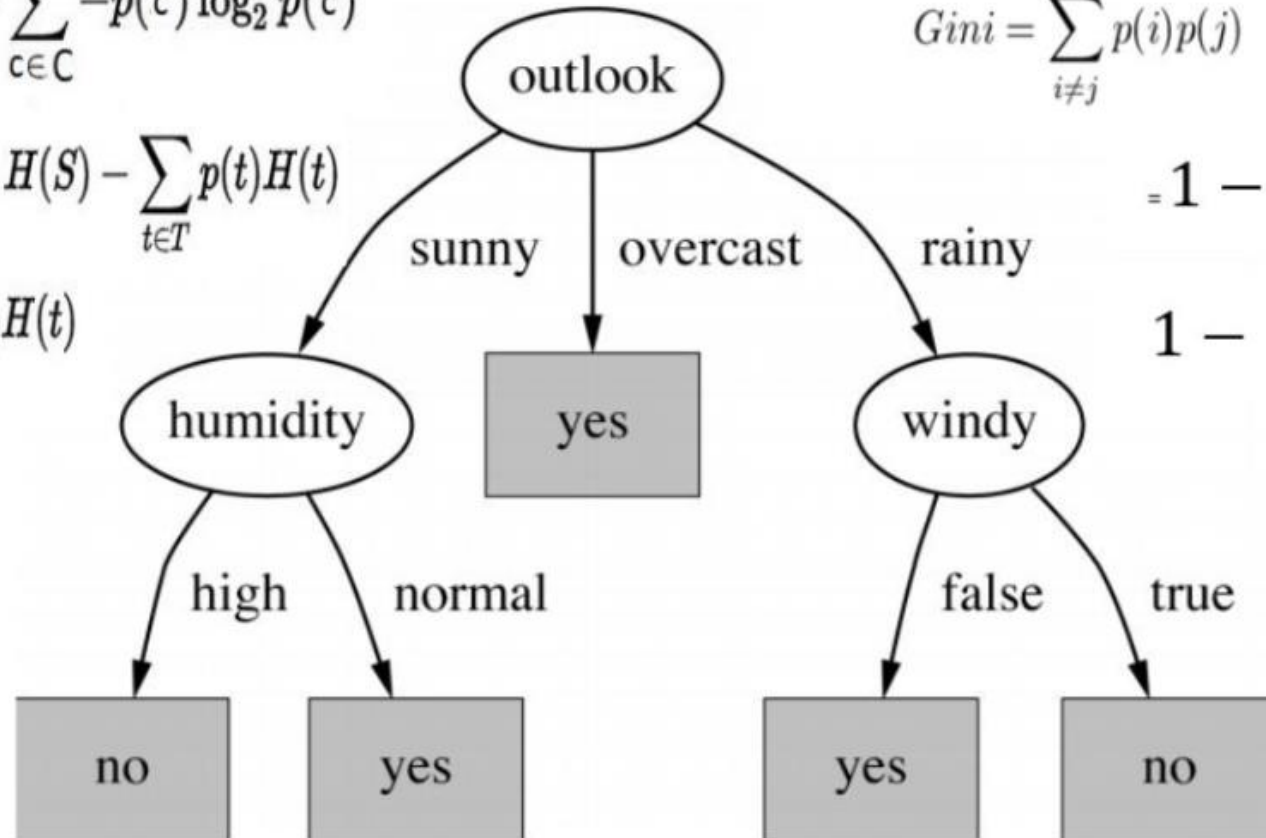
$$\sum_{t \in T} p(t)H(t)$$

GINI Index

$$Gini = \sum_{i \neq j} p(i)p(j)$$

$$= 1 - \sum_{t=0}^{t=1} P_t^2$$

$$1 - \sum_{t=0}^{t=k} P_t^2$$

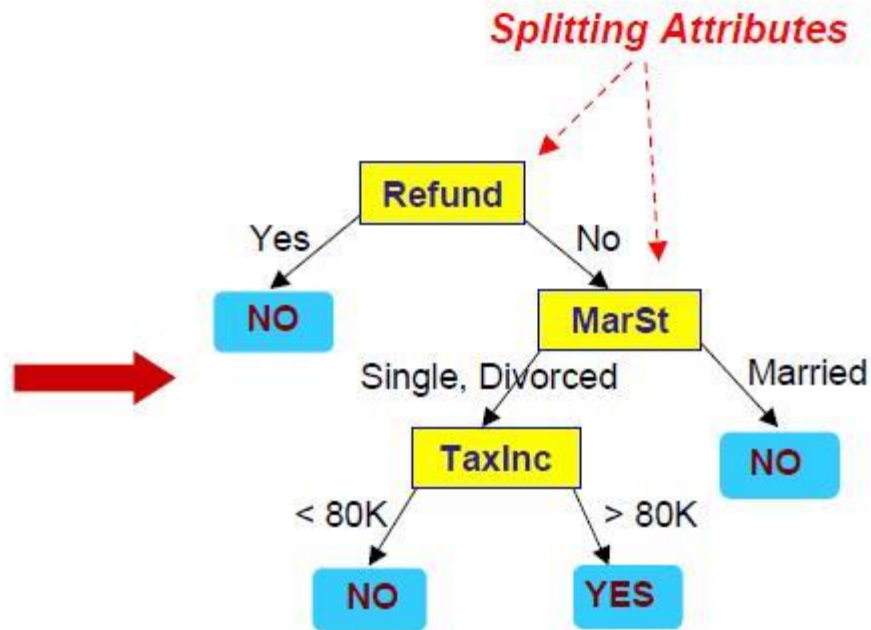


Decision Tree

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data



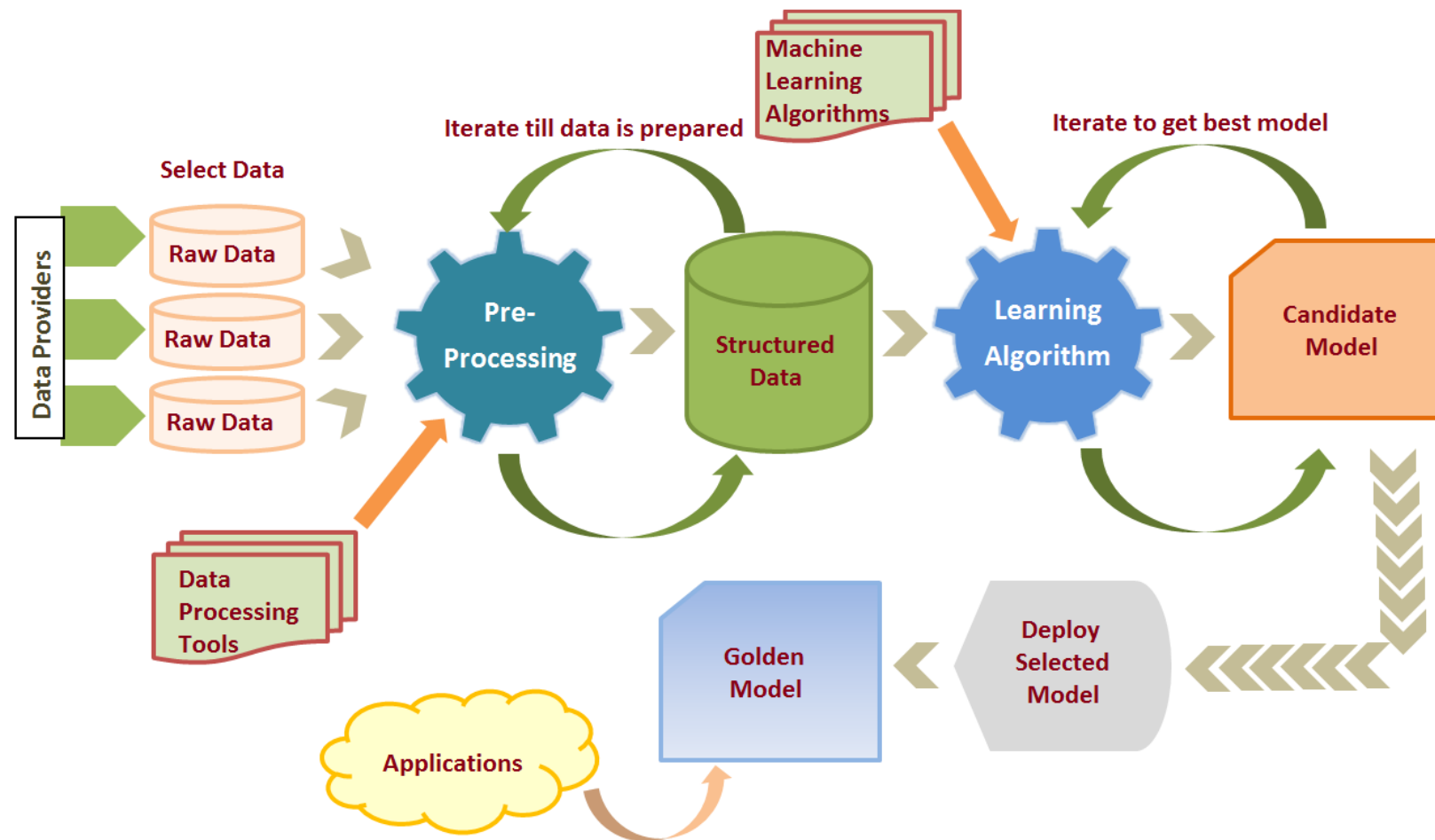
Model: Decision Tree

Naïve Bayes

- Naive Bayes is a probabilistic classifier inspired by the Bayes theorem under a simple assumption which is the attributes are conditionally independent.

$$P(\mathbf{X} | C_i) = \prod_{k=1}^n P(x_k | C_i) = P(x_1 | C_i) \times P(x_2 | C_i) \times \dots \times P(x_n | C_i)$$

3. Methodology



Dataset

ลำดับที่	หมู่	ตำบล	อำเภอ	จังหวัด	พื้นที่ปลูก (ไร่)	พันธุ์ข้าว	วันที่ปลูก	วันที่เก็บเกี่ยว	แหล่งน้ำ	ผลผลิต
1	14	บึงทองหลาง	ลำลูกกา	ปทุมธานี	9.50	กข57	20160730	20161117	มีแหล่งน้ำ	6,624.00
2	14	บึงทองหลาง	ลำลูกกา	ปทุมธานี	24.13	กข57	20160802	20161120	มีแหล่งน้ำ	17,280.00
3	13	คลองสี	คลองหลวง	ปทุมธานี	1.50	กข41	20160515	20160905	มีแหล่งน้ำ	1,300.00
4	1	บึงคอไห	ลำลูกกา	ปทุมธานี	11.03	กข31 (ปท)	20160728	20161119	มีแหล่งน้ำ	10,980.00
5	1	บึงคอไห	ลำลูกกา	ปทุมธานี	14.00	กข47	20160724	20161115	มีแหล่งน้ำ	10,248.00
6	7	คลองควา	สามโคก	ปทุมธานี	20.00	กข41	20160501	20160901	มีแหล่งน้ำ	18,000.00
7	1	ลำลูกกา	ลำลูกกา	ปทุมธานี	15.00	กข31 (ปท)	20160703	20161020	มีแหล่งน้ำ	11,205.00
8	2	พืชอุดม	ลำลูกกา	ปทุมธานี	30.00	กข41	20160610	20161010	มีแหล่งน้ำ	26,352.00
9	4	คลองสอง	คลองหลวง	ปทุมธานี	18.00	กข31 (ปท)	20160514	20160905	มีแหล่งน้ำ	15,000.00
10	2	ลำลูกกา	ลำลูกกา	ปทุมธานี	25.00	กข31 (ปท)	20160609	20160930	มีแหล่งน้ำ	20,000.00
11	5	ระแหง	ลาดหลุมแก้ว	ปทุมธานี	9.50	กข41	20160721	20161110	มีแหล่งน้ำ	9,000.00
12	5	คลองควา	สามโคก	ปทุมธานี	11.00	กข57	20160725	20161104	มีแหล่งน้ำ	8,437.00
13	3	คลองควา	สามโคก	ปทุมธานี	5.00	กข31 (ปท)	20160821	20161130	มีแหล่งน้ำ	4,000.00
14	13	คลองสี	คลองหลวง	ปทุมธานี	6.00	กข41	20160603	20160920	มีแหล่งน้ำ	5,100.00
15	14	บึงทองหลาง	ลำลูกกา	ปทุมธานี	12.52	ปทุมธานี 1	20160727	20161114	มีแหล่งน้ำ	8,784.00

Preparing data


transform data

ลำดับที่	หมู่	ตำบล	อำเภอ	จังหวัด	พื้นที่ปลูก (ไร่)	พันธุ์ข้าว	วันที่ปลูก	วันที่เก็บเกี่ยว	แหล่งน้ำ	ผลผลิต
1	14	บึงทองหล	ลำลูกกา	ปทุมธานี	9.50	กข57	20160730	20161117	มีแหล่งน้ำ	6,624.00
2	14	บึงทองหล	ลำลูกกา	ปทุมธานี	24.13	กข57	20160802	20161120	มีแหล่งน้ำ	17,280.00
3	13	คลองสี	คลองหลวง	ปทุมธานี	1.50	กข41	20160515	20160905	มีแหล่งน้ำ	1,300.00
4	1	บึงคอไห	ลำลูกกา	ปทุมธานี	11.03	กข31 (ปท	20160728	20161119	มีแหล่งน้ำ	10,980.00
5	1	บึงคอไห	ลำลูกกา	ปทุมธานี	14.00	กข47	20160724	20161115	มีแหล่งน้ำ	10,248.00
6	7	คลองควา	สามโคก	ปทุมธานี	20.00	กข41	20160501	20160901	มีแหล่งน้ำ	18,000.00
7	1	ลำลูกกา	ลำลูกกา	ปทุมธานี	15.00	กข31 (ปท	20160703	20161020	มีแหล่งน้ำ	11,205.00
8	2	พิชอุดม	ลำลูกกา	ปทุมธานี	30.00	กข41	20160610	20161010	มีแหล่งน้ำ	26,352.00
9	4	คลองสอง	คลองหลวง	ปทุมธานี	18.00	กข31 (ปท	20160514	20160905	มีแหล่งน้ำ	15,000.00
10	2	ลำลูกกา	ลำลูกกา	ปทุมธานี	25.00	กข31 (ปท	20160609	20160930	มีแหล่งน้ำ	20,000.00



ลำดับที่	หมู่	ตำบล	อำเภอ	จังหวัด	พื้นที่ปลูก (ไร่)	พันธุ์ข้าว	วันที่ปลูก	วันที่เก็บเกี่ยว	ระยะเวลา	แหล่งน้ำ	ผลผลิต
1	14	บึงทองหล	ลำลูกกา	ปทุมธานี	9.50	กข57	7/30/2016	11/17/2016	110	มีแหล่งน้ำ	6,624.00
2	14	บึงทองหล	ลำลูกกา	ปทุมธานี	24.13	กข57	8/2/2016	11/20/2016	110	มีแหล่งน้ำ	17,280.00
3	13	คลองสี	คลองหลวง	ปทุมธานี	1.50	กข41	5/15/2016	9/5/2016	113	มีแหล่งน้ำ	1,300.00
4	1	บึงคอไห	ลำลูกกา	ปทุมธานี	11.03	กข31 (ปท	7/28/2016	11/19/2016	114	มีแหล่งน้ำ	10,980.00
5	1	บึงคอไห	ลำลูกกา	ปทุมธานี	14.00	กข47	7/24/2016	11/15/2016	114	มีแหล่งน้ำ	10,248.00
6	7	คลองควา	สามโคก	ปทุมธานี	20.00	กข41	5/1/2016	9/1/2016	123	มีแหล่งน้ำ	18,000.00
7	1	ลำลูกกา	ลำลูกกา	ปทุมธานี	15.00	กข31 (ปท	7/3/2016	10/20/2016	109	มีแหล่งน้ำ	11,205.00
8	2	พิชอุดม	ลำลูกกา	ปทุมธานี	30.00	กข41	6/10/2016	10/10/2016	122	มีแหล่งน้ำ	26,352.00
9	4	คลองสอง	คลองหลวง	ปทุมธานี	18.00	กข31 (ปท	5/14/2016	9/5/2016	114	มีแหล่งน้ำ	15,000.00
10	2	ลำลูกกา	ลำลูกกา	ปทุมธานี	25.00	กข31 (ปท	6/9/2016	9/30/2016	113	มีแหล่งน้ำ	20,000.00

Preparing data (Cont)



ลำดับที่	หมู่	ตำบล	อำเภอ	จังหวัด	พื้นที่ปลูก (ไร่)	พันธุ์ข้าว	วันที่ปลูก	วันที่เก็บเกี่ยว	ระยะเวลา	แหล่งน้ำ	ผลผลิต	ค่าเฉลี่ยต่อไร่
1	14	บึงทองหล	ลำนูกกา	ปทุมธานี	9.50	กข57	7/30/2016	11/17/2016	110	มีแหล่งน้ำ	6,624.00	697.263158
2	14	บึงทองหล	ลำนูกกา	ปทุมธานี	24.13	กข57	8/2/2016	11/20/2016	110	มีแหล่งน้ำ	17,280.00	716.046825
3	13	คลองสี	คลองหลวง	ปทุมธานี	1.50	กข41	5/15/2016	9/5/2016	113	มีแหล่งน้ำ	1,300.00	866.666667
4	1	บึงคอไห	ลำนูกกา	ปทุมธานี	11.03	กข31 (ปท	7/28/2016	11/19/2016	114	มีแหล่งน้ำ	10,980.00	995.692587
5	1	บึงคอไห	ลำนูกกา	ปทุมธานี	14.00	กข47	7/24/2016	11/15/2016	114	มีแหล่งน้ำ	10,248.00	732
6	7	คลองควา	สามโคก	ปทุมธานี	20.00	กข41	5/1/2016	9/1/2016	123	มีแหล่งน้ำ	18,000.00	900
7	1	ลำนูกกา	ลำนูกกา	ปทุมธานี	15.00	กข31 (ปท	7/3/2016	10/20/2016	109	มีแหล่งน้ำ	11,205.00	747
8	2	พีชอุดม	ลำนูกกา	ปทุมธานี	30.00	กข41	6/10/2016	10/10/2016	122	มีแหล่งน้ำ	26,352.00	878.4
9	4	คลองสอง	คลองหลวง	ปทุมธานี	18.00	กข31 (ปท	5/14/2016	9/5/2016	114	มีแหล่งน้ำ	15,000.00	833.333333
10	2	ลำนูกกา	ลำนูกกา	ปทุมธานี	25.00	กข31 (ปท	6/9/2016	9/30/2016	113	มีแหล่งน้ำ	20,000.00	800

Preparing data (Cont)

Feature selection

ลำดับที่	หมู่	ตำบล	อำเภอ	จังหวัด	พื้นที่ปลูก (ไร่)	พันธุ์ข้าว	วันที่ปลูก	วันที่เก็บเกี่ยว	ระยะเวลา	แหล่งน้ำ	ผลผลิต	ค่าเฉลี่ยต่อไร่
1	14	บึงทองหล	ลำลูกกา	ปทุมธานี	9.50	กข57	7/30/2016	11/17/2016	110	มีแหล่งน้ำ	6,624.00	697.263158
2	14	บึงทองหล	ลำลูกกา	ปทุมธานี	24.13	กข57	8/2/2016	11/20/2016	110	มีแหล่งน้ำ	17,280.00	716.046825
3	13	คลองสี	คลองหลวง	ปทุมธานี	1.50	กข41	5/15/2016	9/5/2016	113	มีแหล่งน้ำ	1,300.00	866.666667
4	1	บึงคอไห	ลำลูกกา	ปทุมธานี	11.03	กข31 (ปท)	7/28/2016	11/19/2016	114	มีแหล่งน้ำ	10,980.00	995.692587
5	1	บึงคอไห	ลำลูกกา	ปทุมธานี	14.00	กข47	7/24/2016	11/15/2016	114	มีแหล่งน้ำ	10,248.00	732
6	7	คลองควา	สามโคก	ปทุมธานี	20.00	กข41	5/1/2016	9/1/2016	123	มีแหล่งน้ำ	18,000.00	900
7	1	ลำลูกกา	ลำลูกกา	ปทุมธานี	15.00	กข31 (ปท)	7/3/2016	10/20/2016	109	มีแหล่งน้ำ	11,205.00	747
8	2	พีชอุดม	ลำลูกกา	ปทุมธานี	30.00	กข41	6/10/2016	10/10/2016	122	มีแหล่งน้ำ	26,352.00	878.4
9	4	คลองสอง	คลองหลวง	ปทุมธานี	18.00	กข31 (ปท)	5/14/2016	9/5/2016	114	มีแหล่งน้ำ	15,000.00	833.333333
10	2	ลำลูกกา	ลำลูกกา	ปทุมธานี	25.00	กข31 (ปท)	6/9/2016	9/30/2016	113	มีแหล่งน้ำ	20,000.00	800

Preprocess (Cont)

Data Transformation

ตำบล (D-1 ถึง D-42)

เขียงรากใหญ่	D-1
คลองควาย	D-2
คลองพระอุดม	D-3
คลองหนึ่ง	D-4
คลองสอง	D-5

อำเภอ (C-1 ถึง C-6)

คลองหลวง	C-1
ฉะบุรี	C-2
ลาดหลุมแก้ว	C-3
ลำลูกกา	C-4
สามโคก	C-5
หนองเสือ	C-6

พันธุ์ข้าว (R-1 ถึง R-20)

กข31 (ปทุมธานี 80 , ซีอีโอ)	R-1
กข37	R-2
กข39	R-3
กข41	R-4
กข47	R-5
กข49	R-6
กข51	R-7

ผลผลิตต่อไร่

LOW	413-609
MIDDLE	610-806
HIGH	807-1000

Preprocess (Cont)

Postprocessing Result

Tambun	District	Area	'rice varie	Duration	range
D-24	C-4	9.5	R-8	110	MIDDLE
D-24	C-4	24.13	R-8	110	MIDDLE
D-7	C-1	1.5	R-4	113	HIGH
D-21	C-4	11.03	R-1	114	HIGH
D-21	C-4	14	R-5	114	MIDDLE
D-2	C-5	20	R-4	123	HIGH
D-37	C-4	15	R-1	109	MIDDLE
D-30	C-4	30	R-4	122	HIGH
D-5	C-1	18	R-1	114	HIGH

Preparing data (Cont)

Dataset for data mining process

ตำบล	อำเภอ	พื้นที่ปลูก (ไร่)	พันธุ์ข้าว	ระยะเวลา	ค่าเฉลี่ยต่อไร่
บึงทองหล	ลำลูกกา	9.50	กข57	110	697.263158
บึงทองหล	ลำลูกกา	24.13	กข57	110	716.046825
คลองสี	คลองหลวง	1.50	กข41	113	866.666667
บึงคอไห	ลำลูกกา	11.03	กข31 (ปท	114	995.692587
บึงคอไห	ลำลูกกา	14.00	กข47	114	732
คลองควา	สามโคก	20.00	กข41	123	900
ลำลูกกา	ลำลูกกา	15.00	กข31 (ปท	109	747
พืชอุดม	ลำลูกกา	30.00	กข41	122	878.4
คลองสอง	คลองหลวง	18.00	กข31 (ปท	114	833.333333
ลำลูกกา	ลำลูกกา	25.00	กข31 (ปท	113	800

Area size, type of rice, water quantity, fertilizer, drug kill, type of soil

Data Model

The image displays two overlapping windows from the Weka software suite. The 'Weka GUI Chooser' window on the left features a bird logo and the text 'WEKA The University of Waikato' and 'Waikato Environment for Knowledge Analysis Version 3.6.11 (c) 1999 - 2014 The University of Waikato Hamilton, New Zealand'. The 'Weka Explorer' window on the right shows the 'cpu' relation with 8 attributes. The 'Selected attribute' section displays a table for the 'vendor' attribute:

No.	Label	Count
1	adviser	1
2	amdahl	9
3	apollo	2
4	basf	2
5	bti	2
6	burroughs	8
7	c.r.d	6

Below this table is a bar chart showing the distribution of the 'vendor' attribute. The 'Class: class (Num)' section shows a distribution with a peak count of 32. The 'Tambun' attribute is also shown with a table and a bar chart. The 'Tambun' table is as follows:

No.	Label	Count
1	D-24	100
2	D-7	45
3	D-21	23
4	D-2	16
5	D-37	52
6	D-30	19
7	D-5	5

The 'Tambun' bar chart shows a distribution with a peak count of 100. The status bar at the bottom of the Weka Explorer window indicates 'Status OK'.

Algorithms



- J48 or C4.5
- NBTree
- RandomTree
- Naïve Bayes

4. Experimental Result

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

When True Positive (TP)

True Negative (TN)

False Positive (FP)

False Negative (FN)

Split 80-20

Algorithms	Accuracy
------------	----------

J48	81 %
-----	------

NBTree	78.5 %
--------	--------

RandomTree	76 %
------------	------

NB	75%
----	-----

Use training set

Algorithms	Accuracy
------------	----------

J48	83.6 %
-----	--------

NBTree	82.8 %
--------	--------

RandomTree	99.3 %
------------	--------

NB	82.4%
----	-------

Cross-Validation Folds 10

Algorithms	Accuracy
------------	----------

J48	78.4 %
-----	--------

NBTree	78.8 %
--------	--------

RandomTree	69.5 %
------------	--------

NB	72.5 %
----	--------

Average

Algorithms	Accuracy
------------	----------

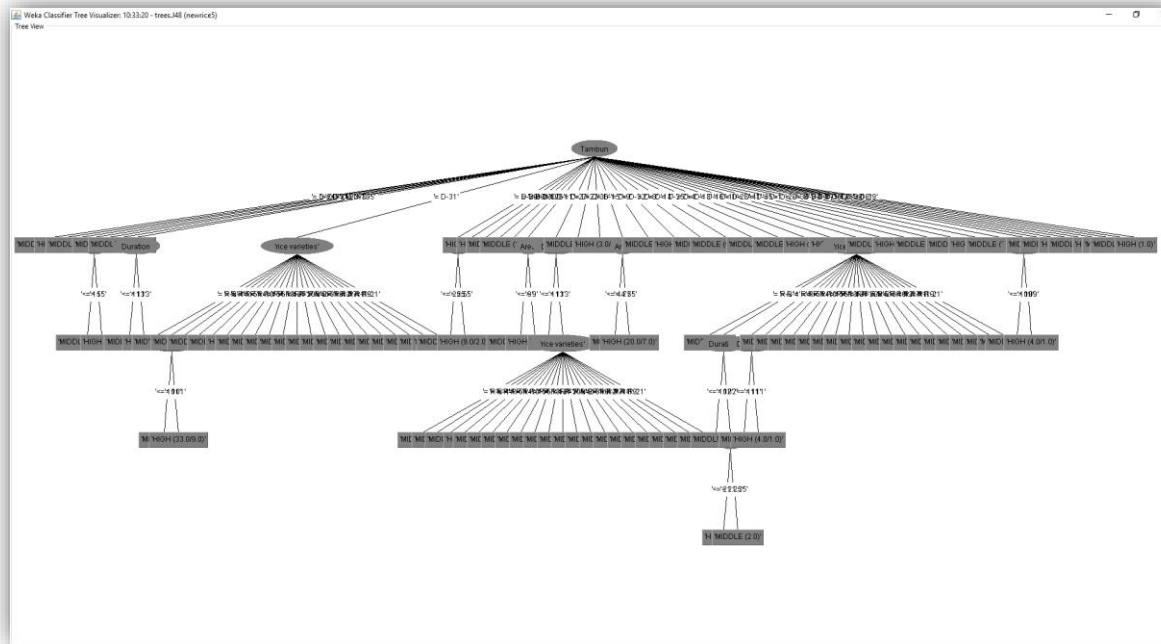
J48	81 %
-----	------

NBTree	80 %
--------	------

RandomTree	81.6 %
------------	--------

NB	76.63 %
----	---------

Rice Prediction Model

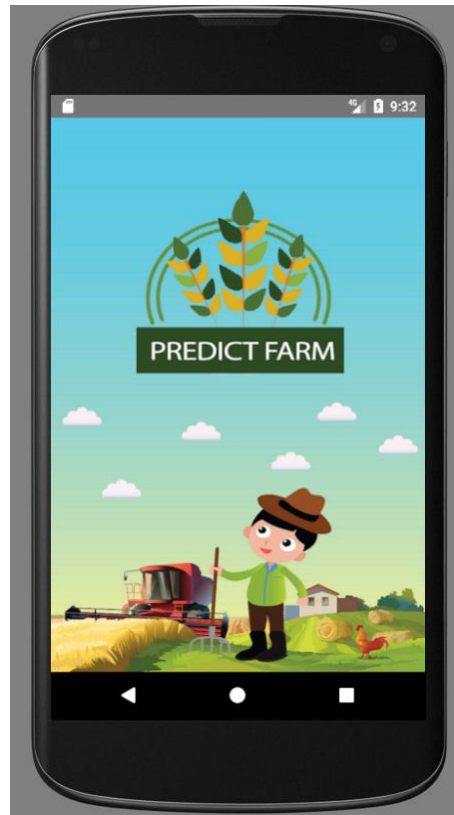


Data Model

```
J48 pruned tree
-----
Tambun = D-24: MIDDLE (100.0/14.0)
Tambun = D-7: HIGH (45.0/12.0)
Tambun = D-21: MIDDLE (23.0/6.0)
Tambun = D-2
| Area <= 15: MIDDLE (8.0/4.0)
| Area > 15: HIGH (8.0)
Tambun = D-37: MIDDLE (52.0/12.0)
Tambun = D-30: MIDDLE (19.0/6.0)
Tambun = D-5
| Duration <= 113: MIDDLE (2.0)
| Duration > 113: HIGH (3.0)
Tambun = D-31
| rice varieties = R-8: MIDDLE (2.0)
| rice varieties = R-4
| | Duration <= 101: MIDDLE (3.0)
| | Duration > 101: HIGH (33.0/9.0)
| rice varieties = R-1: MIDDLE (30.0/9.0)
| rice varieties = R-5: MIDDLE (2.0/1.0)
| rice varieties = R-14: MIDDLE (0.0)
| rice varieties = R-10: HIGH (3.0)
| rice varieties = R-15: MIDDLE (1.0)
| rice varieties = R-6: MIDDLE (0.0)
| rice varieties = R-13: MIDDLE (1.0)
| rice varieties = R-18: MIDDLE (0.0)
| rice varieties = R-7: MIDDLE (1.0)
| rice varieties = R-20 : MIDDLE (0.0)
```

Model Rules

5.Application



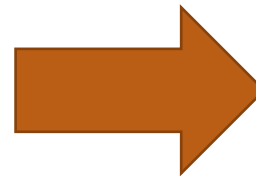
Home Menu



Main Functions

Application (Cont)

Prediction Form

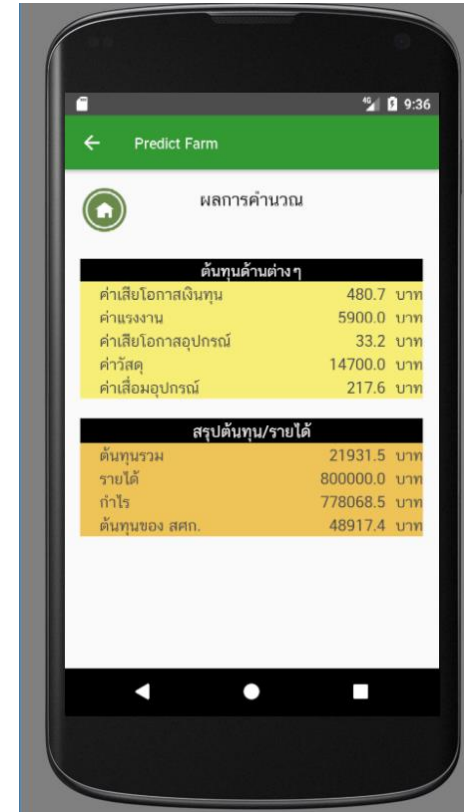
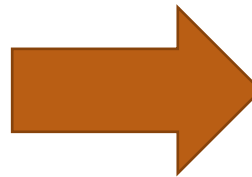


Result of Prediction Form

Application (Cont)



Cost Prediction

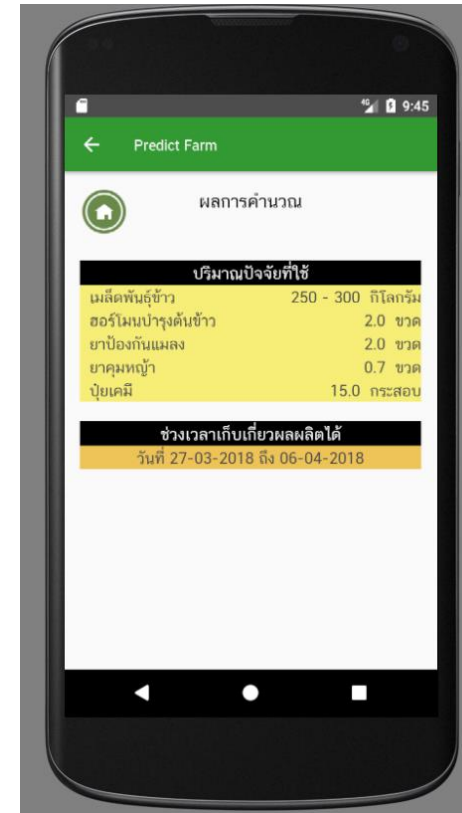
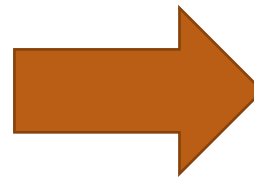


Result of Cost Prediction

Application (Cont)

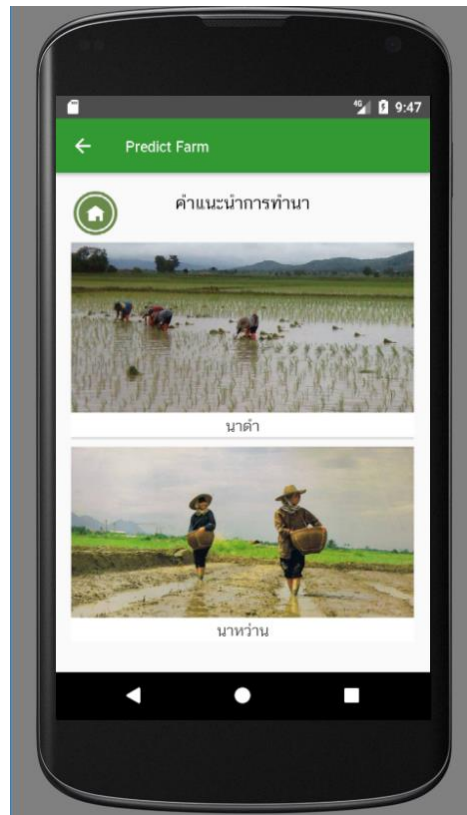


Factor Calculation



Result of Factor Calculation

Application (Cont)



Recommender Form

6. Application Assessment



