

## 1.1 載入 train.arff，將 PassengerId、Name、Ticket 欄位刪除

使用 remove filter 刪除這三個欄位：

The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. The 'Filter' section shows 'Remove -R 1,4,9' applied. The 'Current relation' section shows 'Relation: train-weka.filters.unsupervised.attribute.Remove...' and 'Instances: 891'. The 'Attributes' section shows a list of attributes with checkboxes: 1 ☐ Survived, 2 ☐ Pclass, 3 ☐ Sex, 4 ☐ Age, 5 ☐ SibSp, 6 ☐ Parch, 7 ☐ Fare, 8 ☐ Cabin, 9 ☐ Embarked. The 'Selected attribute' section shows 'Name: Survived' with statistics: Minimum 0, Maximum 1, Mean 0.384, StdDev 0.487. The 'Class: Embarked (Nom)' section shows a bar chart with three bars: 0 (blue, 549), 1 (red, 342), and 2 (cyan, 0). The status bar shows 'OK' and a 'Log' button.

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choo... | Remove -R 1,4,9 | Apply | Stop

Current relation

Relation: train-weka.filters.unsupervised.attribute.Remove... | Attributes: 9 | Sum of weights: 891

Instances: 891

Attributes

All | None | Invert | Pattern

No.	Name
1	<input type="checkbox"/> Survived
2	<input type="checkbox"/> Pclass
3	<input type="checkbox"/> Sex
4	<input type="checkbox"/> Age
5	<input type="checkbox"/> SibSp
6	<input type="checkbox"/> Parch
7	<input type="checkbox"/> Fare
8	<input type="checkbox"/> Cabin
9	<input type="checkbox"/> Embarked

Remove

Selected attribute

Name: Survived | Type: Numeric | Missing: 0 (0%) | Distinct: 2 | Unique: 0 (0%)

Statistic	Value
Minimum	0
Maximum	1
Mean	0.384
StdDev	0.487

Class: Embarked (Nom) | Visualize All

549 | 342 | 0

0 | 0.5 | 1

Status: OK | Log | x 0

## 1.2 將 Cabin 的非空值以 1 替代，空值以 0 填入

先用 MergeManyValues 取代非空值，填入1:

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choo... MergeManyValues -C 8 -L 1 -R first-last

Apply

Stop

Current relation

Relation: train-weka.filters.unsupervised.attribute.MergeManyValues

Instances: 891

Attributes

All

None

No.	
1	<input type="checkbox"/> Survived
2	<input type="checkbox"/> Pclass
3	<input type="checkbox"/> Sex
4	<input type="checkbox"/> Age
5	<input type="checkbox"/> SibSp
6	<input type="checkbox"/> Parch
7	<input type="checkbox"/> Fare
8	<input type="checkbox"/> Cabin
9	<input type="checkbox"/> Embarked

Remove

Status

OK

Selected attribute

Type: Numeric

Unique: 97 (11%)

Value

0

512.329

32.204

49.693

Visualize All

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.MergeManyValues

About

Merges many values of a nominal attribute into one value.

More

Capabilities

attributeIndex

8

debug

False

doNotCheckCapabilities

False

ignoreClass

False

label

1

mergeValueRange

first-last

Open...

Save...

OK

Cancel

53

39

47

8

11

7

9

2

0

4

5

2

6

0

0

0

0

0

0

0

0

0

0

3

0

256.16

512.33

Log

x 0

填入1後的結果:

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Viewer


Relation: train-weka.filters.unsupervised.attribute.Remove-R1,4,9-weka.filters.unsupervised.attribute.MergeManyValues-C8-L1-Rfirst-last

No.	1: Survived Numeric	2: Pclass Numeric	3: Sex Nominal	4: Age Numeric	5: SibSp Numeric	6: Parch Numeric	7: Fare Numeric	8: Cabin Nominal	9: Embarked Nominal
1	0.0	3.0	male	22.0	1.0	0.0	7.25		S
2	1.0	1.0	female	38.0	1.0	0.0	71.2833	1	C
3	1.0	3.0	female	26.0	0.0	0.0	7.925		S
4	1.0	1.0	female	35.0	1.0	0.0	53.1	1	S
5	0.0	3.0	male	35.0	0.0	0.0	8.05		S
6	0.0	3.0	male		0.0	0.0	8.4583		Q
7	0.0	1.0	male	54.0	0.0	0.0	51.8625	1	S
8	0.0	3.0	male	2.0	3.0	1.0	21.075		S
9	1.0	3.0	female	27.0	0.0	2.0	11.1333		S
10	1.0	2.0	female	14.0	1.0	0.0	30.0708		C
11	1.0	3.0	female	4.0	1.0	1.0	16.7	1	S
12	1.0	1.0	female	58.0	0.0	0.0	26.55	1	S
13	0.0	3.0	male	20.0	0.0	0.0	8.05		S
14	0.0	3.0	male	39.0	1.0	5.0	31.275		S
15	0.0	3.0	female	14.0	0.0	0.0	7.8542		S
16	1.0	2.0	female	55.0	0.0	0.0	16.0		S
17	0.0	3.0	male	2.0	4.0	1.0	29.125		Q
18	1.0	2.0	male		0.0	0.0	13.0		S
19	0.0	3.0	female	31.0	1.0	0.0	18.0		S
20	1.0	3.0	female		0.0	0.0	7.225		C
21	0.0	2.0	male	35.0	0.0	0.0	26.0		S
22	1.0	2.0	male	34.0	0.0	0.0	13.0	1	S
23	1.0	3.0	female	15.0	0.0	0.0	8.0292		Q
24	1.0	1.0	male	28.0	0.0	0.0	35.5	1	S

Add instanceUndoOKCancel

Remove

Status  
OK

Log x 0

再用 ReplaceMissingWithUserConstant 取代空值，填入0:

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choo... ReplaceMissingWithUserConstant -A 8 -N 0 -R 0 -F "yyyy-MM-dd'T\HH:mm:ss"

Apply

Stop

Current relation

Relation: train-weka.filters.unsupervised.att

Instances: 891

Attributes

All

None

No.

1

☐ Survived

2

☐ Pclass

3

☐ Sex

4

☐ Age

5

☐ SibSp

6

☐ Parch

7

☐ Fare

8

☐ Cabin

9

☐ Embarked

Remove

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.ReplaceMissingWithUserConstant

About

Replaces all missing values for nominal, string, numeric and date attributes in the dataset with user-supplied constant values.

attributes

8

dateFormat

yyyy-MM-dd'T\HH:mm:ss

dateReplacementValue

debug

False

doNotCheckCapabilities

False

ignoreClass

False

nominalStringReplacementValue

0

numericReplacementValue

0

Open...

Save...

OK

Cancel

More

Capabilities

Type: Nominal

Unique: 0 (0%)

Count

Weight

1

204

Visualize All

Status

OK

Log

x 0

填入0後的完整結果:

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Viewer

Relation: train-weka.filters.unsupervised.attribute.Remove-R1,4,9-weka.filters.unsupervised.attribute.MergeManyValues-C8-L1-Rfirst-last-weka.filters.unsupervised.attribute.R...

No.	1: Survived Numeric	2: Pclass Numeric	3: Sex Nominal	4: Age Numeric	5: SibSp Numeric	6: Parch Numeric	7: Fare Numeric	8: Cabin Nominal	9: Embarked Nominal
1	0.0	3.0	male	22.0	1.0	0.0	7.25	0	S
2	1.0	1.0	female	38.0	1.0	0.0	71.2833	1	C
3	1.0	3.0	female	26.0	0.0	0.0	7.925	0	S
4	1.0	1.0	female	35.0	1.0	0.0	53.1	1	S
5	0.0	3.0	male	35.0	0.0	0.0	8.05	0	S
6	0.0	3.0	male		0.0	0.0	8.4583	0	Q
7	0.0	1.0	male	54.0	0.0	0.0	51.8625	1	S
8	0.0	3.0	male	2.0	3.0	1.0	21.075	0	S
9	1.0	3.0	female	27.0	0.0	2.0	11.1333	0	S
10	1.0	2.0	female	14.0	1.0	0.0	30.0708	0	C
11	1.0	3.0	female	4.0	1.0	1.0	16.7	1	S
12	1.0	1.0	female	58.0	0.0	0.0	26.55	1	S
13	0.0	3.0	male	20.0	0.0	0.0	8.05	0	S
14	0.0	3.0	male	39.0	1.0	5.0	31.275	0	S
15	0.0	3.0	female	14.0	0.0	0.0	7.8542	0	S
16	1.0	2.0	female	55.0	0.0	0.0	16.0	0	S
17	0.0	3.0	male	2.0	4.0	1.0	29.125	0	Q
18	1.0	2.0	male		0.0	0.0	13.0	0	S
19	0.0	3.0	female	31.0	1.0	0.0	18.0	0	S
20	1.0	3.0	female		0.0	0.0	7.225	0	C
21	0.0	2.0	male	35.0	0.0	0.0	26.0	0	S
22	1.0	2.0	male	34.0	0.0	0.0	13.0	1	S
23	1.0	3.0	female	15.0	0.0	0.0	8.0292	0	Q
24	1.0	1.0	male	28.0	0.0	0.0	35.5	1	S

Remove

Add instance

Undo

OK

Cancel

Status

OK

Log

x 0

1.3 使用 ReplaceMissingValues 將 Age 的空值以 Age 平均數填入

將 Age 的空值以 Age 平均數填入：

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

Filter

Choo...ReplaceMissingValuesApplyStop

Current relation

Relation: train-weka.filters.unsupervised.attribute.Remove...  
Instances: 891

Attributes: 9  
Sum of weights: 891

Selected attribute

Name: Survived  
Missing: 0 (0%)  
Distinct: 2

Type: Nominal  
Unique: 0 (0%)

Attributes

AllNone

No.	
1	<input type="checkbox"/> Survived
2	<input type="checkbox"/> Pclass
3	<input type="checkbox"/> Sex
4	<input type="checkbox"/> Age
5	<input type="checkbox"/> SibSp
6	<input type="checkbox"/> Parch
7	<input type="checkbox"/> Fare
8	<input type="checkbox"/> Cabin
9	<input type="checkbox"/> Embarked

Remove

No.	Label	Count	Weight
		549	
		342	

Visualize All

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.ReplaceMissingValues

About

Replaces all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data.

MoreCapabilities

debugFalse

doNotCheckCapabilitiesFalse

ignoreClassFalse

Open...Save...OKCancel

Status

OK

Log

x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Viewer

Relation: train-weka.filters.unsupervised.attribute.Remove-R1,4,9-weka.filters.unsupervised.attribute.MergeManyValues-C8-L1-Rfirst-last-weka.filters.unsupervised.attribute.R...

No.	1: Survived Numeric	2: Pclass Numeric	3: Sex Nominal	4: Age Numeric	5: SibSp Numeric	6: Parch Numeric	7: Fare Numeric	8: Cabin Nominal	9: Embarked Nominal
1	0.0	3.0	male	22.0	1.0	0.0	7.25	0	S
2	1.0	1.0	female	38.0	1.0	0.0	71.2833	1	C
3	1.0	3.0	female	26.0	0.0	0.0	7.925	0	S
4	1.0	1.0	female	35.0	1.0	0.0	53.1	1	S
5	0.0	3.0	male	35.0	0.0	0.0	8.05	0	S
6	0.0	3.0	male	29.69...	0.0	0.0	8.4583	0	Q
7	0.0	1.0	male	54.0	0.0	0.0	51.8625	1	S
8	0.0	3.0	male	2.0	3.0	1.0	21.075	0	S
9	1.0	3.0	female	27.0	0.0	2.0	11.1333	0	S
10	1.0	2.0	female	14.0	1.0	0.0	30.0708	0	C
11	1.0	3.0	female	4.0	1.0	1.0	16.7	1	S
12	1.0	1.0	female	58.0	0.0	0.0	26.55	1	S
13	0.0	3.0	male	20.0	0.0	0.0	8.05	0	S
14	0.0	3.0	male	39.0	1.0	5.0	31.275	0	S
15	0.0	3.0	female	14.0	0.0	0.0	7.8542	0	S
16	1.0	2.0	female	55.0	0.0	0.0	16.0	0	S
17	0.0	3.0	male	2.0	4.0	1.0	29.125	0	Q
18	1.0	2.0	male	29.69...	0.0	0.0	13.0	0	S
19	0.0	3.0	female	31.0	1.0	0.0	18.0	0	S
20	1.0	3.0	female	29.69...	0.0	0.0	7.225	0	C
21	0.0	2.0	male	35.0	0.0	0.0	26.0	0	S
22	1.0	2.0	male	34.0	0.0	0.0	13.0	1	S
23	1.0	3.0	female	15.0	0.0	0.0	8.0292	0	Q
24	1.0	1.0	male	28.0	0.0	0.0	35.5	1	S

Remove

Add instance Undo OK Cancel

Status OK

Log

使用 filter 的 NumericToNominal 將 Survived 與 Pclass 轉為 Nominal:

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

Filter

Choo...NumericToNominal -R 1,2ApplyStop

Current relation

Relation: train-weka.filters.unsupervised.attribute.Remove...Attributes: 9Sum of weights: 891

Selected attribute

Name: SurvivedMissing: 0 (0%)Distinct: 2Type: NumericUnique: 0 (0%)

Attributes

AllNone

No.	
1	<input type="checkbox"/> Survived
2	<input type="checkbox"/> Pclass
3	<input type="checkbox"/> Sex
4	<input type="checkbox"/> Age
5	<input type="checkbox"/> SibSp
6	<input type="checkbox"/> Parch
7	<input type="checkbox"/> Fare
8	<input type="checkbox"/> Cabin
9	<input type="checkbox"/> Embarked

Remove

Status

OK

Log

x 0

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.NumericToNominal

About

A filter for turning numeric attributes into nominal ones.

MoreCapabilities

attributeIndices1,2

debugFalse

doNotCheckCapabilitiesFalse

invertSelectionFalse

Open...Save...OKCancel

Value

0

1

0.384

0.487

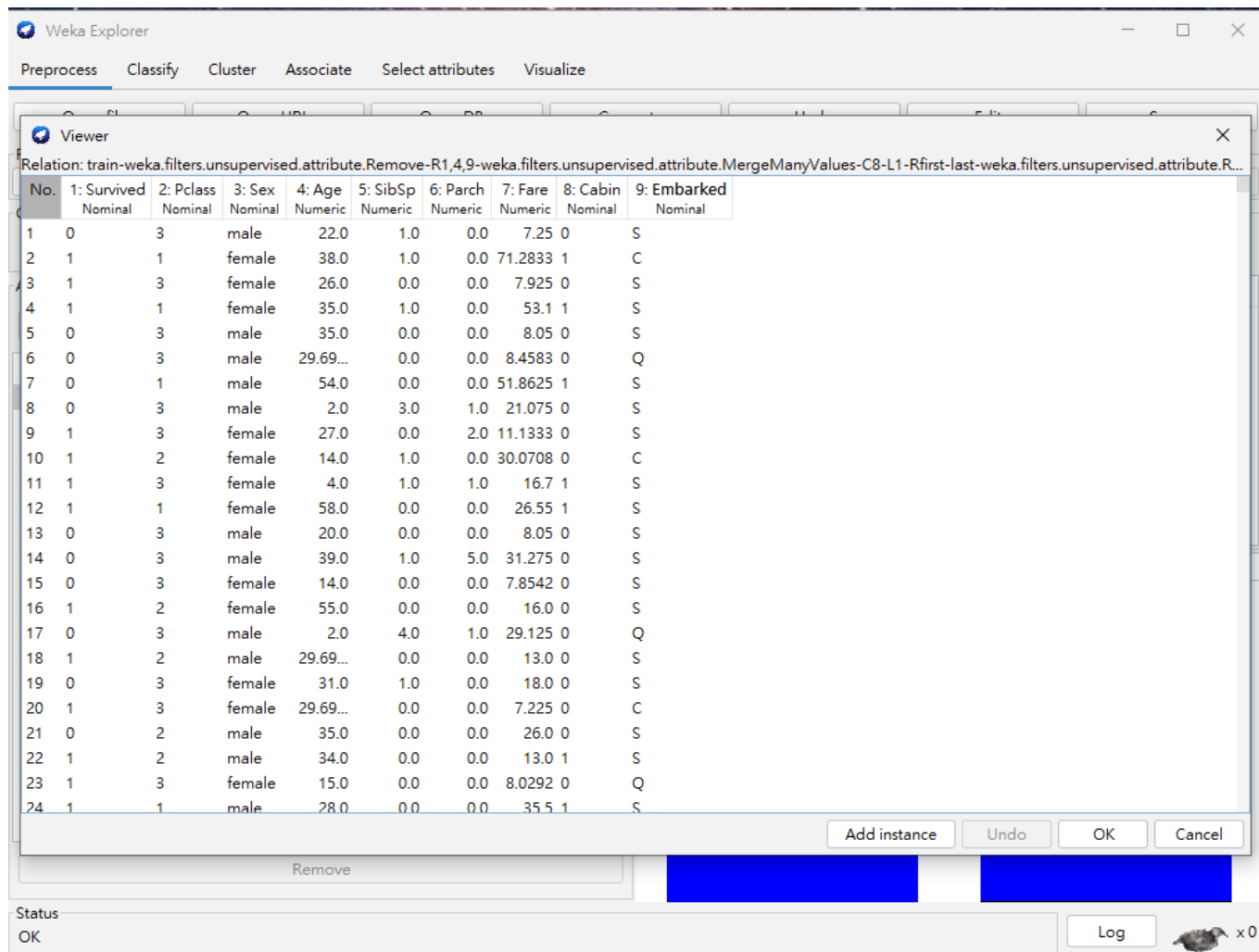
Visualize All

342

00.51



轉換之後結果如下:



Relation: train-weka.filters.unsupervised.attribute.Remove-R1,4,9-weka.filters.unsupervised.attribute.MergeManyValues-C8-L1-Rfirst-last-weka.filters.unsupervised.attribute.R...

No.	1: Survived Nominal	2: Pclass Nominal	3: Sex Nominal	4: Age Numeric	5: SibSp Numeric	6: Parch Numeric	7: Fare Numeric	8: Cabin Nominal	9: Embarked Nominal
1	0	3	male	22.0	1.0	0.0	7.25	0	S
2	1	1	female	38.0	1.0	0.0	71.2833	1	C
3	1	3	female	26.0	0.0	0.0	7.925	0	S
4	1	1	female	35.0	1.0	0.0	53.1	1	S
5	0	3	male	35.0	0.0	0.0	8.05	0	S
6	0	3	male	29.69...	0.0	0.0	8.4583	0	Q
7	0	1	male	54.0	0.0	0.0	51.8625	1	S
8	0	3	male	2.0	3.0	1.0	21.075	0	S
9	1	3	female	27.0	0.0	2.0	11.1333	0	S
10	1	2	female	14.0	1.0	0.0	30.0708	0	C
11	1	3	female	4.0	1.0	1.0	16.7	1	S
12	1	1	female	58.0	0.0	0.0	26.55	1	S
13	0	3	male	20.0	0.0	0.0	8.05	0	S
14	0	3	male	39.0	1.0	5.0	31.275	0	S
15	0	3	female	14.0	0.0	0.0	7.8542	0	S
16	1	2	female	55.0	0.0	0.0	16.0	0	S
17	0	3	male	2.0	4.0	1.0	29.125	0	Q
18	1	2	male	29.69...	0.0	0.0	13.0	0	S
19	0	3	female	31.0	1.0	0.0	18.0	0	S
20	1	3	female	29.69...	0.0	0.0	7.225	0	C
21	0	2	male	35.0	0.0	0.0	26.0	0	S
22	1	2	male	34.0	0.0	0.0	13.0	1	S
23	1	3	female	15.0	0.0	0.0	8.0292	0	Q
24	1	1	male	28.0	0.0	0.0	35.5	1	S

Buttons: Add instance, Undo, OK, Cancel

Status: OK

Log x 0

為何 **Numeric** 無法使用在 **Decision tree**?

要用在 **Decision tree** 的數值必須先經過離散化，故連續的 **Numeric** 無法被用在 **Decision tree**，要先做離散化處理轉換為 **Nominal**

**1.5 以 70% 切割訓練資料，使用 J48 對 Survived 進行分類**

## 分類準確率、混淆矩陣:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: J48 -C 0.25 -M 2

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds: 10
- ☒ Percentage split %: 70

(Nom) Survived

Result list (right-click for options)

- 23:53:18 - trees.J48
- 23:55:11 - trees.J48

Classifier output

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	218	81.6479 %
Incorrectly Classified Instances	49	18.3521 %
Kappa statistic	0.5769	
Mean absolute error	0.266	
Root mean squared error	0.3719	
Relative absolute error	56.4711 %	
Root relative squared error	77.0588 %	
Total Number of Instances	267	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.947	0.408	0.800	0.947	0.867	0.599	0.820	0.833	0
	0.592	0.053	0.866	0.592	0.703	0.599	0.820	0.762	1
Weighted Avg.	0.816	0.278	0.824	0.816	0.807	0.599	0.820	0.807	

=== Confusion Matrix ===

a	b	-- classified as
160	9	a = 0
40	58	b = 1

Status: OK

x 0

視覺化的Decision tree:

