

Predictions in(CSV format ) the Predicted A16 column

FinalPredit.csv - Excel																		Amila	
Tell me what you want to do...																			
Q20																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q		
1	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	'prediction margin	'predicted A16'		
2	b	32.67	y	p	9	w	0	FALSE	h	5.25	TRUE	0	TRUE	154	g	0.503932	Success		
3	?	20.08	u	g	0.125	q	768	TRUE	v	1	FALSE	1	FALSE	240	g	-0.87395	Failure		
4	b	20.08	u	g	0.25	q	0	FALSE	v	0.125	FALSE	0	FALSE	200	g	-0.87395	Failure		
5	b	22.17	u	g	2.25	i	10	FALSE	v	0.125	FALSE	0	FALSE	160	g	-0.87395	Failure		
6	a	27.25	u	g	0.29	m	108	TRUE	h	0.125	FALSE	1	TRUE	272	g	-0.87395	Failure		
7	b	31.58	y	p	0.75	aa	0	FALSE	v	3.5	FALSE	0	TRUE	320	g	-0.87395	Failure		
8	a	20.83	u	g	8.5	c	351	FALSE	v	0.165	FALSE	0	FALSE	0	g	-0.87395	Failure		
9	b	48.08	u	g	3.75	i	2	FALSE	bb	1	FALSE	0	FALSE	100	g	-0.87395	Failure		
10	b	29.83	u	g	3.5	c	0	FALSE	v	0.165	FALSE	0	FALSE	216	g	-0.87395	Failure		
11	a	41.58	u	g	1.04	aa	237	FALSE	v	0.665	FALSE	0	FALSE	240	g	-0.87395	Failure		
12	b	33.17	u	g	1.04	r	31285	FALSE	h	6.5	TRUE	0	TRUE	164	g	0.503932	Success		
13	a	18.92	u	g	9	aa	591	TRUE	v	0.75	TRUE	2	FALSE	88	g	0.79	Success		
14	a	24.75	u	g	3	q	500	TRUE	h	1.835	TRUE	19	FALSE	0	g	0.79	Success		
15	b	21	y	p	4.79	w	300	TRUE	v	2.25	TRUE	1	TRUE	80	g	0.79	Success		
16																			

## ARFF format of the predicted data

```
C:\Users\PLUS VISION PC\Desktop\Semester 6\CO 544 - Machine Learning and Data Mining\Project\2\final1aa.arff - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
config.js test1result1 data.csv.arff testdata_10%.csv.arff final1.arff final2.arff final1aa.arff
1 @relation 'testdata_10%-weka.filters.unsupervised.attribute.StringToNominal-Rlast_predicted'
2
3 @attribute A1 {b,a}
4 @attribute A2 numeric
5 @attribute A3 {u,y,l}
6 @attribute A4 {g,p,gg}
7 @attribute A5 numeric
8 @attribute A6 {w,q,c,x,i,d,e,aa,cc,ff,m,k,j,r}
9 @attribute A7 numeric
10 @attribute A8 {TRUE,FALSE}
11 @attribute A9 {v,h,bb,ff,j,z,o,dd,n}
12 @attribute A10 numeric
13 @attribute A11 {TRUE,FALSE}
14 @attribute A12 numeric
15 @attribute A13 {FALSE,TRUE}
16 @attribute A14 numeric
17 @attribute A15 {g,s,p}
18 @attribute 'prediction margin' numeric
19 @attribute 'predicted A16' {Success,Failure}
20 @attribute A16 {Success,Failure}
21
22 @data
23 b,32.67,y,p,9,w,0,FALSE,h,5.25,TRUE,0,TRUE,154,g,0.503932,Success,?
24 ?,20.08,u,g,0.125,q,768,TRUE,v,1,FALSE,1,FALSE,240,g,-0.87395,Failure,?
25 b,20.08,u,g,0.25,q,0,FALSE,v,0.125,FALSE,0,FALSE,200,g,-0.87395,Failure,?
26 b,22.17,u,g,2.25,i,10,FALSE,v,0.125,FALSE,0,FALSE,160,g,-0.87395,Failure,?
27 a,27.25,u,g,0.29,m,108,TRUE,h,0.125,FALSE,1,TRUE,272,g,-0.87395,Failure,?
28 b,31.58,y,p,0.75,aa,0,FALSE,v,3.5,FALSE,0,TRUE,320,g,-0.87395,Failure,?
29 a,20.83,u,g,8.5,c,351,FALSE,v,0.165,FALSE,0,FALSE,0,g,-0.87395,Failure,?
30 b,48.08,u,g,3.75,i,2,FALSE,bb,1,FALSE,0,FALSE,100,g,-0.87395,Failure,?
31 b,29.83,u,g,3.5,c,0,FALSE,v,0.165,FALSE,0,FALSE,216,g,-0.87395,Failure,?
32 a,41.58,u,g,1.04,aa,237,FALSE,v,0.665,FALSE,0,FALSE,240,g,-0.87395,Failure,?
33 b,33.17,u,g,1.04,r,31285,FALSE,h,6.5,TRUE,0,TRUE,164,g,0.503932,Success,?
34 a,18.92,u,g,9,aa,591,TRUE,v,0.75,TRUE,2,FALSE,88,g,0.79,Success,?
35 a,24.75,u,g,3,q,500,TRUE,h,1.835,TRUE,19,FALSE,0,g,0.79,Success,?
36 b,21,y,p,4.79,w,300,TRUE,v,2.25,TRUE,1,TRUE,80,g,0.79,Success,?
```

Normal text file length : 1,607 lines : 36

defined as the difference between the probability predicted for the actual class and the highest probability predicted for the other classes. A margin of 1 means that the correct class is predicted with 100% confidence (very good), a margin of -1 means that an incorrect class is predicted with 100% confidence (very bad).

1).Both Training and Test data sets were preprocessed and made to ARFF formats. For test data sets Attribute names were given and the Class (Which is to be predicted is names as "A16")

Figure 1: Modified Test data set

```

C:\Users\PLUS VISION PC\Desktop\Semester 6\CO 544 - Machine Learning and Data Mining\Project\2\testdata_10%.arff - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
config.js test1result1 data.csv.arff testdata_10%.arff final1.arff final2.arff final1aa.arff
1 @relation 'testdata_10%-weka.filters.unsupervised.attribute.StringToNominal-R1ast'
2
3 @attribute A1 {b,a}
4 @attribute A2 numeric
5 @attribute A3 {u,y,l}
6 @attribute A4 {g,p,gg}
7 @attribute A5 numeric
8 @attribute A6 {w,q,c,x,i,d,e,aa,cc,ff,m,k,j,r}
9 @attribute A7 numeric
10 @attribute A8 {TRUE,FALSE}
11 @attribute A9 {v,h,bb,ff,j,z,o,dd,n}
12 @attribute A10 numeric
13 @attribute A11 {TRUE,FALSE}
14 @attribute A12 numeric
15 @attribute A13 {FALSE,TRUE}
16 @attribute A14 numeric
17 @attribute A15 {g,s,p}
18 @attribute A16 {Success,Failure}
19
20 @data
21 b,32.67,y,p,9,w,0,FALSE,h,5.25,TRUE,0,TRUE,154,g,?
22 ?,20.08,u,g,0.125,q,768,TRUE,v,1,FALSE,1,FALSE,240,g,?
23 b,20.08,u,g,0.25,q,0,FALSE,v,0.125,FALSE,0,FALSE,200,g,?
24 b,22.17,u,g,2.25,i,10,FALSE,v,0.125,FALSE,0,FALSE,160,g,?
25 a,27.25,u,g,0.29,m,108,TRUE,h,0.125,FALSE,1,TRUE,272,g,?
26 b,31.58,y,p,0.75,aa,0,FALSE,v,3.5,FALSE,0,TRUE,320,g,?
27 a,20.83,u,g,8.5,c,351,FALSE,v,0.165,FALSE,0,FALSE,0,g,?
28 b,48.08,u,g,3.75,i,2,FALSE,bb,1,FALSE,0,FALSE,100,g,?
29 b,29.83,u,g,3.5,c,0,FALSE,v,0.165,FALSE,0,FALSE,216,g,?
30 a,41.58,u,g,1.04,aa,237,FALSE,v,0.665,FALSE,0,FALSE,240,g,?
31 b,33.17,u,g,1.04,x,31285,FALSE,h,6.5,TRUE,0,TRUE,164,g,?
32 a,18.92,u,g,9,aa,591,TRUE,v,0.75,TRUE,2,FALSE,88,g,?
33 a,24.75,u,g,3,q,500,TRUE,h,1.835,TRUE,19,FALSE,0,g,?
34 b,21,y,p,4.75,w,300,TRUE,v,2.25,TRUE,1,TRUE,80,g,?
35
36

```

Figure 2: Modified test data in ARFF format

## 2) Training Data set classified Using J48 tree(Use as training data set)

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

☒ Use training set

☐ Supplied test set Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) A16

Start Stop

Result list (right-click for options)

17:15:49 - trees.J48

Classifier output

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	501	90.7609 %
Incorrectly Classified Instances	51	9.2391 %
Kappa statistic	0.8147	
Mean absolute error	0.1629	
Root mean squared error	0.2846	
Relative absolute error	32.7754 %	
Root relative squared error	57.0763 %	
Total Number of Instances	552	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.922	0.104	0.883	0.922	0.902	0.815	0.929	0.888	Success
	0.896	0.078	0.930	0.896	0.913	0.815	0.929	0.914	Failure
Weighted Avg.	0.908	0.090	0.909	0.908	0.908	0.815	0.929	0.902	

=== Confusion Matrix ===

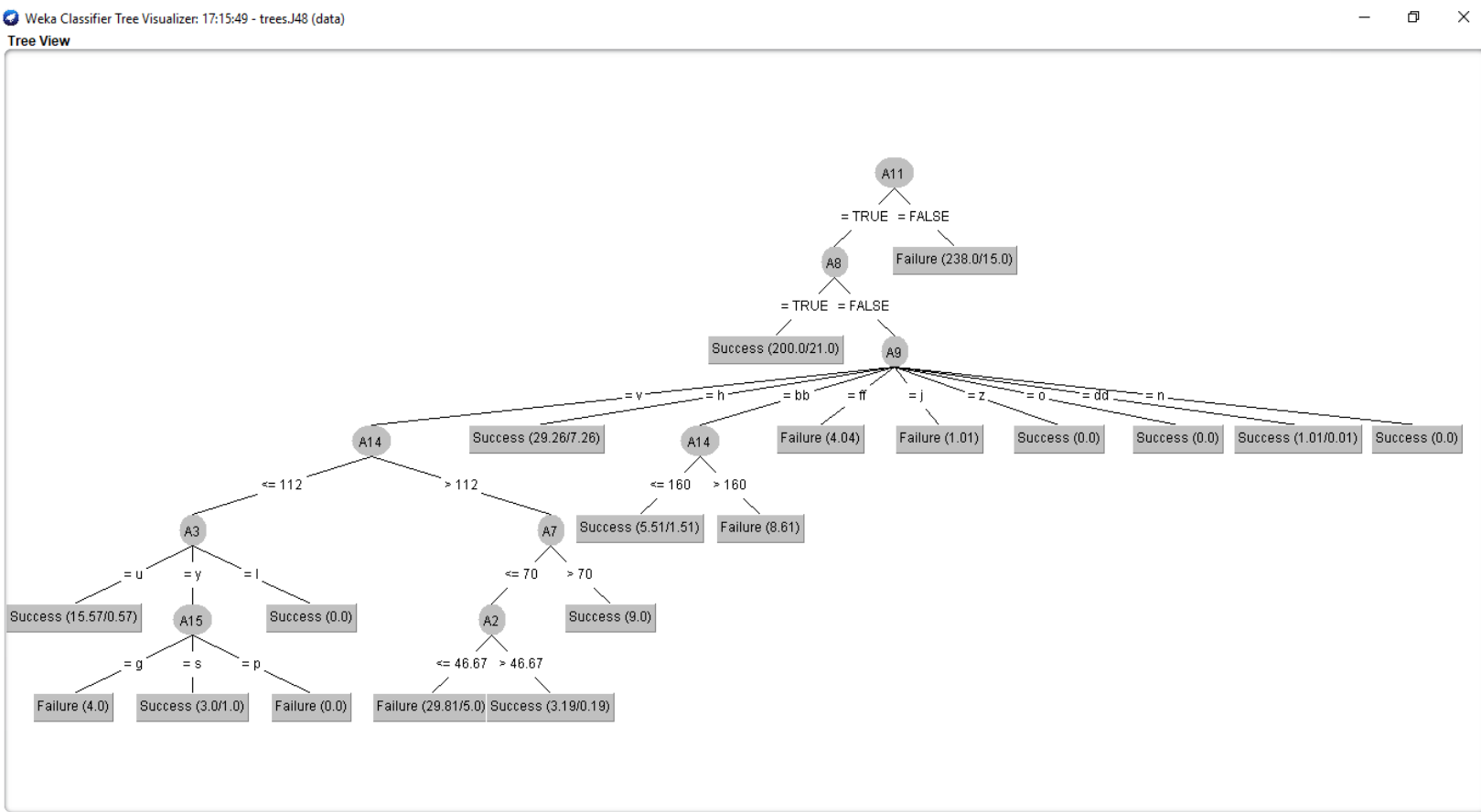
```

a b <-- classified as
235 20 | a = Success
31 266 | b = Failure

```

**Correctly Classified 90.7609%**

Visualised tree(for training data set)



**3) Supplied the test data set and classified Using J48 algorithm.**

**Results were saved.**

## Results

FinalPredit.csv - Excel

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do...

Q20

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	'prediction margin	'predicted A16
2	b	32.67	y	p	9	w	0	FALSE	h	5.25	TRUE	0	TRUE	154	g	0.503932	Success
3	?	20.08	u	g	0.125	q	768	TRUE	v	1	FALSE	1	FALSE	240	g	-0.87395	Failure
4	b	20.08	u	g	0.25	q	0	FALSE	v	0.125	FALSE	0	FALSE	200	g	-0.87395	Failure
5	b	22.17	u	g	2.25	i	10	FALSE	v	0.125	FALSE	0	FALSE	160	g	-0.87395	Failure
6	a	27.25	u	g	0.29	m	108	TRUE	h	0.125	FALSE	1	TRUE	272	g	-0.87395	Failure
7	b	31.58	y	p	0.75	aa	0	FALSE	v	3.5	FALSE	0	TRUE	320	g	-0.87395	Failure
8	a	20.83	u	g	8.5	c	351	FALSE	v	0.165	FALSE	0	FALSE	0	g	-0.87395	Failure
9	b	48.08	u	g	3.75	i	2	FALSE	bb	1	FALSE	0	FALSE	100	g	-0.87395	Failure
10	b	29.83	u	g	3.5	c	0	FALSE	v	0.165	FALSE	0	FALSE	216	g	-0.87395	Failure
11	a	41.58	u	g	1.04	aa	237	FALSE	v	0.665	FALSE	0	FALSE	240	g	-0.87395	Failure
12	b	33.17	u	g	1.04	r	31285	FALSE	h	6.5	TRUE	0	TRUE	164	g	0.503932	Success
13	a	18.92	u	g	9	aa	591	TRUE	v	0.75	TRUE	2	FALSE	88	g	0.79	Success
14	a	24.75	u	g	3	q	500	TRUE	h	1.835	TRUE	19	FALSE	0	g	0.79	Success
15	b	21	y	p	4.79	w	300	TRUE	v	2.25	TRUE	1	TRUE	80	g	0.79	Success
16																	
17																	

Figure 3: Final Predictions.(Column labeled as “Predicted A16”)

## Conclusion

**Prediction margin:** This is defined as the difference between the probability predicted for the actual class and the highest probability predicted for the other classes. A margin of 1 means that the correct class is predicted with 100% confidence (very good), a margin of -1 means that an incorrect class is predicted with 100% confidence (very bad).