

Project 2: Data Pre-processing, Mining, and Evaluation of
Classification Rules: Homework

Piotr Mardziel

7th April 2004

Contents

2 Project 2: Data Pre-processing, Mining, and Evaluation of Classification Rules: Homework	2
2.1 Group Members	2
2.2 Pre-Project Homework Assignment	2
2.2.1 PRISM Using $\frac{P}{t}$	2
2.2.1.1 risk = low	2
2.2.1.2 risk = moderate	5
2.2.1.3 risk = high	9
2.2.1.4 Results	13
2.2.2 PRISM Using $p \cdot [\lg(\frac{P}{t}) - \lg(\frac{P}{T})]$	13
2.2.2.1 risk = low	13
2.2.2.2 risk = moderate	16
2.2.2.3 risk = high	19
2.2.2.4 Results	23
2.2.3 PRISM Pruning	23

Chapter 2

Project 2: Data Pre-processing, Mining, and Evaluation of Classification Rules: Homework

2.1 Group Members

GROUP MEMBER	USERNAME	TASKS
Piotr Mardziel	piotrm	everything

2.2 Pre-Project Homework Assignment

2.2.1 PRISM Using $\frac{p}{t}$

Lets start with rules for determining high risk.

2.2.1.1 risk = low

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.1.1.1 Rule 1

- IF ? THEN risk = low

Candidate Attributes	
Candidate	Goodness ²
credit_history = bad	0/4
credit_history = unknown	2/5
credit_history = good	3/5
dept = low	3/7
debt = high	2/7
collateral = none	3/11
collateral = adequate	2/3
income = 0-15	0/3
income = 15-35	0/5
income = >35	5/6

- IF income = >35 AND ? THEN risk = low

Rule is not perfect so lets try again given only those instances covered by this rule so far.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
13	good	high	none	>35	low

Candidate Attributes	
Candidate	Goodness
credit_history = bad	0/1
credit_history = unknown	2/2
credit_history = good	3/3
dept = low	3/4
debt = high	2/2
collateral = none	3/3
collateral = adequate	2/3

- IF income = >35 AND credit_history = good THEN risk = low

The rule is now perfect so lets remove the instances covered and start over to get another rule.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate

Instance	credit_history	debt	collateral	income	risk
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

2.2.1.1.2 Rule 2

- IF ? THEN risk = low

Candidate Attributes	
Candidate	Goodness
credit_history = bad	0/4
credit_history = unknown	2/5
credit_history = good	0/5
dept = low	2/6
debt = high	0/5
collateral = none	1/9
collateral = adequate	1/2
income = 0-15	0/3
income = 15-35	0/5
income = >35	2/3

- IF income = >35 THEN risk = low

The rule is not perfect. So lets keep on going.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate

Candidate Attributes	
Candidate	Goodness
credit_history = bad	0/1
credit_history = unknown	2/2
credit_history = good	0/0
dept = low	2/3
debt = high	0/0
collateral = none	1/1
collateral = adequate	1/2

- IF income = >35 AND credit_history = unknown THEN risk = low

This rule is perfect. Remove the instances it classifies and continue.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

There are no more instances with risk = low so this part of the rule-set is complete. The two rules generated are:

2.2.1.1.3 Resulting Rules

1. IF income = >35 AND credit_history = good THEN risk = low
2. IF income = >35 AND credit_history = unknown THEN risk = low

Next comes moderate risk.

2.2.1.2 risk = moderate

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.1.2.1 Rule 1

- IF ? THEN risk = moderate

Candidate Attributes	
Candidate	Goodness
credit_history = bad	2/4
credit_history = unknown	1/5
credit_history = good	1/5
dept = low	3/7

Candidate	Goodness
debt = high	1/7
collateral = none	3/11
collateral = adequate	1/3
income = 0-15	0/3
income = 15-35	3/5
income = >35	1/6

- IF income = 15-35 THEN risk = moderate

This rule is not even close to being perfect so more attributes are necessary.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

- IF income = 15-35 AND ? THEN risk = moderate

Candidate Attributes	
Candidate	Goodness
credit_history = bad	1/2
credit_history = unknown	1/2
credit_history = good	1/1
debt = low	2/2
debt = high	1/3
collateral = none	3/5
collateral = adequate	0/0

- IF income = 15-35 AND debt = low THEN risk = moderate

This rule is perfect. Removing the instances classified and trying to get another rule.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.1.2.2 Rule 2

- IF ? THEN risk = moderate

Candidate Attributes	
Candidate	Goodness
credit_history = bad	1/3
credit_history = unknown	0/4
credit_history = good	1/5
dept = low	1/5
debt = high	1/7
collateral = none	1/9
collateral = adequate	1/3
income = 0-15	0/3
income = 15-35	1/3
income = >35	1/6

- IF credit_history = bad THEN risk = moderate

This seems to be the least perfect partial rule so far so let's keep on going.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
8	bad	low	adequate	>35	moderate
14	bad	high	none	15-35	high

- IF credit_history = bad AND ? THEN risk = moderate

Candidate Attributes	
Candidate	Goodness
dept = low	1/2
debt = high	0/1
collateral = none	0/2
collateral = adequate	1/1
income = 0-15	0/1
income = 15-35	0/1
income = >35	1/1

- IF credit_history = bad AND collateral = adequate THEN risk = moderate

This rule is perfect so lets remove the instances it classifies and try to get a third rule.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low

Instance	credit_history	debt	collateral	income	risk
7	unknown	high	none	0-15	high
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.1.2.3 Rule 3

- IF ? THEN risk = moderate

Candidate Attributes	
Candidate	Goodness
credit_history = bad	0/2
credit_history = unknown	0/4
credit_history = good	1/5
dept = low	0/4
debt = high	1/7
collateral = none	1/9
collateral = adequate	0/2
income = 0-15	0/3
income = 15-35	1/3
income = >35	0/5

- IF income = 15-35 THEN risk = moderate

Unfortunately the rule is not yet perfect so lets continue.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

- IF income = 15-35 AND ? THEN risk = moderate

Candidate Attributes	
Candidate	Goodness
credit_history = bad	0/1
credit_history = unknown	0/1
credit_history = good	1/1
dept = low	0/0
debt = high	1/3
collateral = none	1/3
collateral = adequate	0/0

- IF income = 15-35 AND credit_history = good THEN risk = moderate

This rule is perfect so let's see if there are any attributes left to classify for a fourth rule.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
13	good	high	none	>35	low
14	bad	high	none	15-35	high

There doesn't appear to be any. This means the risk = moderate rules are done.

2.2.1.2.4 Resulting Rules

1. IF income = 15-35 AND debt = low THEN risk = moderate
2. IF credit_history = bad AND collateral = adequate THEN risk = moderate
3. IF income = 15-35 AND credit_history = good THEN risk = moderate

Finally approaches high risk.

2.2.1.3 risk = high

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.1.3.1 Rule 1

- IF ? THEN risk = high

Candidate Attributes	
Candidate	Goodness
credit_history = bad	2/4
credit_history = unknown	2/5

Candidate	Goodness
credit_history = good	1/5
dept = low	1/7
debt = high	4/7
collateral = none	5/11
collateral = adequate	0/3
income = 0-15	3/3
income = 15-35	2/5
income = >35	0/6

- **IF income = 0-15 THEN risk = high**

Luckily this rule is already perfect so already the instances classified by it can be removed and rule 2 can be searched for.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.1.3.2 Rule 2

- IF ? THEN risk = high

Candidate Attributes	
Candidate	Goodness
credit_history = bad	1/3
credit_history = unknown	1/4
credit_history = good	0/4
dept = low	0/6
debt = high	2/5
collateral = none	2/8
collateral = adequate	0/3
income = 0-15	0/0
income = 15-35	2/5
income = >35	0/6

- IF debt = high THEN risk = high

Rule is not yet perfect.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
10	good	high	adequate	>35	low
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

- IF debt = high AND ? THEN risk = high

Candidate Attributes	
Candidate	Goodness
credit_history = bad	1/1
credit_history = unknown	1/1
credit_history = good	0/3
collateral = none	2/4
collateral = adequate	0/1
income = 0-15	0/0
income = 15-35	2/3
income = >35	0/2

- **IF debt = high AND credit_history = bad THEN risk = high**

Now the rule is perfect but there is room for one more rule.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
12	good	high	none	15-35	moderate
13	good	high	none	>35	low

2.2.1.3.3 Rule 3

- IF ? THEN risk = high

Candidate Attributes	
Candidate	Goodness
credit_history = bad	0/2
credit_history = unknown	1/4
credit_history = good	0/4
debt = low	0/6
debt = high	1/4

Candidate	Goodness
collateral = none	1/7
collateral = adequate	0/3
income = 0-15	0/0
income = 15-35	1/4
income = >35	0/6

- IF credit_history = unknown THEN risk = high

This is not yet perfect.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low

- IF credit_history = unknown AND ? THEN risk = high

Candidate Attributes	
Candidate	Goodness
dept = low	0/3
debt = high	1/1
collateral = none	1/3
collateral = adequate	0/1
income = 0-15	0/0
income = 15-35	1/2
income = >35	0/2

- IF credit_history = unknown AND debt = high THEN risk = high

Lets check if any more rules are required.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
12	good	high	none	15-35	moderate
13	good	high	none	>35	low

It doesn't look like it as there are no more instances with high risk.

2.2.1.3.4 Resulting Rules

1. IF income = 0-15 THEN risk = high

2. IF debt = high AND credit_history = bad THEN risk = high
3. IF credit_history = unknown AND debt = high THEN risk = high

2.2.1.4 Results

This concludes the computation of PRISM on this dataset. There were 8 rules generated. They are:

1. risk = low
 - (a) IF income = >35 AND credit_history = good THEN risk = low
 - (b) IF income = >35 AND credit_history = unknown THEN risk = low
2. risk = moderate
 - (a) IF income = 15-35 AND debt = low THEN risk = moderate
 - (b) IF credit_history = bad AND collateral = adequate THEN risk = moderate
 - (c) IF income = 15-35 AND credit_history = good THEN risk = moderate
3. risk = high
 - (a) IF income = 0-15 THEN risk = high
 - (b) IF debt = high AND credit_history = bad THEN risk = high
 - (c) IF credit_history = unknown AND debt = high THEN risk = high

2.2.2 PRISM Using $p \cdot \left[\lg \left(\frac{p}{t} \right) - \lg \left(\frac{p}{T} \right) \right]$

2.2.2.1 risk = low

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.2.1.1 Rule 1

- IF ? THEN risk = low $[-\lg \left(\frac{p}{T} \right) = -\lg \left(\frac{5}{14} \right) = 1.485]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$0 \cdot \lg\left(\frac{0}{4}\right) - \lg\left(\frac{5}{14}\right) = -\infty$
credit_history = unknown	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{5}{14}\right) = 0.327$
credit_history = good	$3 \cdot \lg\left(\frac{3}{5}\right) - \lg\left(\frac{5}{14}\right) = 2.245$
dept = low	$3 \cdot \lg\left(\frac{3}{7}\right) - \lg\left(\frac{5}{14}\right) = 0.789$
debt = high	$2 \cdot \lg\left(\frac{2}{7}\right) - \lg\left(\frac{5}{14}\right) = -0.644$
collateral = none	$3 \cdot \lg\left(\frac{3}{11}\right) - \lg\left(\frac{5}{14}\right) = -1.157$
collateral = adequate	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{5}{14}\right) = 1.801$
income = 0-15	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{5}{14}\right) = -\infty$
income = 15-35	$0 \cdot \lg\left(\frac{0}{5}\right) - \lg\left(\frac{5}{14}\right) = -\infty$
income = >35	$5 \cdot \lg\left(\frac{5}{6}\right) - \lg\left(\frac{5}{14}\right) = 6.112$

- IF income = >35 THEN risk = low [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{5}{6}\right) = 0.263$]

The rule is not perfect so lets keep on going.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
13	good	high	none	>35	low

- IF income = >35 AND ? THEN risk = low [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{5}{6}\right) = 0.263$]

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$0 \cdot \lg\left(\frac{0}{7}\right) - \lg\left(\frac{5}{6}\right) = -\infty$
credit_history = unknown	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{5}{6}\right) = 0.526$
credit_history = good	$3 \cdot \lg\left(\frac{3}{3}\right) - \lg\left(\frac{5}{6}\right) = 0.789$
dept = low	$3 \cdot \lg\left(\frac{3}{4}\right) - \lg\left(\frac{5}{6}\right) = -0.456$
debt = high	$2 \cdot \lg\left(\frac{2}{2}\right) - \lg\left(\frac{5}{6}\right) = 0.525$
collateral = none	$3 \cdot \lg\left(\frac{3}{3}\right) - \lg\left(\frac{5}{6}\right) = 0.789$
collateral = adequate	$2 \cdot \lg\left(\frac{2}{3}\right) - \lg\left(\frac{5}{6}\right) = -0.644$

- **IF income = >35 and credit_history = good THEN risk = low** [perfect]

This rule is perfect so lets remove instances it covers and try to find another rule.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low

Instance	credit_history	debt	collateral	income	risk
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

2.2.2.1.2 Rule 2

- IF ? THEN risk = low [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{11}\right) = 2.459$]

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$0 \cdot \lg\left(\frac{0}{4}\right) - \lg\left(\frac{2}{11}\right) = -\infty$
credit_history = unknown	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{2}{11}\right) = 2.275$
credit_history = good	$0 \cdot \lg\left(\frac{0}{7}\right) - \lg\left(\frac{2}{11}\right) = -\infty$
dept = low	$2 \cdot \lg\left(\frac{2}{6}\right) - \lg\left(\frac{2}{11}\right) = 1.750$
debt = high	$0 \cdot \lg\left(\frac{0}{5}\right) - \lg\left(\frac{2}{11}\right) = -\infty$
collateral = none	$1 \cdot \lg\left(\frac{1}{9}\right) - \lg\left(\frac{2}{11}\right) = -0.710$
collateral = adequate	$1 \cdot \lg\left(\frac{1}{2}\right) - \lg\left(\frac{2}{11}\right) = 1.459$
income = 0-15	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{2}{11}\right) = -\infty$
income = 15-35	$0 \cdot \lg\left(\frac{0}{5}\right) - \lg\left(\frac{2}{11}\right) = -\infty$
income = >35	$2 \cdot \lg\left(\frac{2}{3}\right) - \lg\left(\frac{2}{11}\right) = 3.749$

- IF income = >35 THEN risk = low [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{3}\right) = 0.585$]

The rule is obviously not perfect yet.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate

- IF income = >35 AND ? THEN risk = low [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{3}\right) = 0.585$]

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$0 \cdot \lg\left(\frac{0}{1}\right) - \lg\left(\frac{2}{3}\right) = -\infty$
credit_history = unknown	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{2}{3}\right) = 1.170$
credit_history = good	$0 \cdot \lg\left(\frac{0}{0}\right) - \lg\left(\frac{2}{3}\right) = -\infty$
dept = low	$2 \cdot \lg\left(\frac{2}{3}\right) - \lg\left(\frac{2}{3}\right) = 0$
debt = high	$0 \cdot \lg\left(\frac{0}{0}\right) - \lg\left(\frac{2}{3}\right) = -\infty$
collateral = none	$1 \cdot \lg\left(\frac{1}{2}\right) - \lg\left(\frac{2}{3}\right) = -0.415$
collateral = adequate	$1 \cdot \lg\left(\frac{1}{1}\right) - \lg\left(\frac{2}{3}\right) = 0.585$

- IF income = >35 AND credit_history = unknown THEN risk = low [perfect]

This rule is perfect and there are no more unclassified low risk instances.

2.2.2.1.3 Resulting Rules

1. IF income = >35 and credit_history = good THEN risk = low
2. IF income = >35 AND credit_history = unknown THEN risk = low

2.2.2.2 risk = moderate

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.2.2.1 Rule 1

- IF ? THEN risk = moderate $[-\lg(\frac{P}{T}) = -\lg(\frac{4}{14}) = 1.807]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$2 \cdot \lg(\frac{2}{4}) - \lg(\frac{4}{14}) = 1.615$
credit_history = unknown	$1 \cdot \lg(\frac{1}{5}) - \lg(\frac{4}{14}) = -0.516$
credit_history = good	$1 \cdot \lg(\frac{1}{2}) - \lg(\frac{4}{14}) = 0.807$
debt = low	$3 \cdot \lg(\frac{3}{7}) - \lg(\frac{4}{14}) = 1.755$
debt = high	$1 \cdot \lg(\frac{1}{7}) - \lg(\frac{4}{14}) = -1.000$
collateral = none	$3 \cdot \lg(\frac{3}{11}) - \lg(\frac{4}{14}) = -0.201$
collateral = adequate	$1 \cdot \lg(\frac{1}{3}) - \lg(\frac{4}{14}) = 0.222$
income = 0-15	$0 \cdot \lg(\frac{0}{3}) - \lg(\frac{4}{14}) = -\infty$
income = 15-35	$3 \cdot \lg(\frac{3}{5}) - \lg(\frac{4}{14}) = 3.211$
income = >35	$1 \cdot \lg(\frac{1}{6}) - \lg(\frac{4}{14}) = -0.778$

- IF income = 15-35 THEN risk = moderate $[-\lg(\frac{P}{T}) = -\lg(\frac{3}{5}) = 0.737]$

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

- IF income = 15-35 AND ? THEN risk = moderate [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{3}{5}\right) = 0.737$]

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$1 \cdot \lg\left(\frac{1}{2}\right) - \lg\left(\frac{3}{5}\right) = -0.263$
credit_history = unknown	$1 \cdot \lg\left(\frac{1}{2}\right) - \lg\left(\frac{3}{5}\right) = -0.263$
credit_history = good	$1 \cdot \lg\left(\frac{1}{1}\right) - \lg\left(\frac{3}{5}\right) = 0.737$
debt = low	$2 \cdot \lg\left(\frac{2}{2}\right) - \lg\left(\frac{3}{5}\right) = 1.474$
debt = high	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{3}{5}\right) = -0.848$
collateral = none	$3 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{3}{5}\right) = 0$
collateral = adequate	$0 \cdot \lg\left(\frac{0}{0}\right) - \lg\left(\frac{3}{5}\right) = -\infty$

- IF income = 15-35 AND debt = low THEN risk = moderate [perfect]

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.2.2.2 Rule 2

- IF ? THEN risk = moderate [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{12}\right) = 2.585$]

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{2}{12}\right) = 1.000$
credit_history = unknown	$0 \cdot \lg\left(\frac{0}{1}\right) - \lg\left(\frac{2}{12}\right) = -\infty$
credit_history = good	$1 \cdot \lg\left(\frac{1}{5}\right) - \lg\left(\frac{2}{12}\right) = 0.263$
debt = low	$1 \cdot \lg\left(\frac{1}{5}\right) - \lg\left(\frac{2}{12}\right) = 0.263$
debt = high	$1 \cdot \lg\left(\frac{1}{7}\right) - \lg\left(\frac{2}{12}\right) = -0.222$
collateral = none	$1 \cdot \lg\left(\frac{1}{9}\right) - \lg\left(\frac{2}{12}\right) = -0.585$
collateral = adequate	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{2}{12}\right) = 1.000$
income = 0-15	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{2}{12}\right) = -\infty$
income = 15-35	$1 \cdot \lg\left(\frac{1}{6}\right) - \lg\left(\frac{2}{12}\right) = 0$
income = >35	$1 \cdot \lg\left(\frac{1}{6}\right) - \lg\left(\frac{2}{12}\right) = 0$

- IF credit_history = bad THEN risk = moderate [$-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{1}{3}\right) = 1.585$]

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
8	bad	low	adequate	>35	moderate
14	bad	high	none	15-35	high

- IF credit_history = bad AND ? THEN risk = moderate [$-\lg\left(\frac{P}{7}\right) = -\lg\left(\frac{1}{3}\right) = 1.585$]

Candidate Attributes	
Candidate	Goodness
dept = low	$1 \cdot \lg\left(\frac{1}{2}\right) - \lg\left(\frac{1}{3}\right) = 0.58$
debt = high	$0 \cdot \lg\left(\frac{0}{1}\right) - \lg\left(\frac{1}{3}\right) = -\infty$
collateral = none	$0 \cdot \lg\left(\frac{0}{2}\right) - \lg\left(\frac{1}{3}\right) = -\infty$
collateral = adequate	$1 \cdot \lg\left(\frac{1}{1}\right) - \lg\left(\frac{1}{3}\right) = 1.585$
income = 0-15	$0 \cdot \lg\left(\frac{0}{1}\right) - \lg\left(\frac{1}{3}\right) = -\infty$
income = 15-35	$0 \cdot \lg\left(\frac{0}{1}\right) - \lg\left(\frac{1}{3}\right) = -\infty$
income = >35	$1 \cdot \lg\left(\frac{1}{1}\right) - \lg\left(\frac{1}{3}\right) = 1.585$

- IF credit_history = bad AND collateral = adequate THEN risk = moderate [perfect]

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
9	good	low	none	>35	low
10	good	high	adequate	>35	low
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.2.2.3 Rule 3

- IF ? THEN risk = moderate [$-\lg\left(\frac{P}{7}\right) = -\lg\left(\frac{1}{11}\right) = 3.459$]

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$0 \cdot \lg\left(\frac{0}{2}\right) - \lg\left(\frac{1}{11}\right) = -\infty$
credit_history = unknown	$0 \cdot \lg\left(\frac{0}{4}\right) - \lg\left(\frac{1}{11}\right) = -\infty$
credit_history = good	$1 \cdot \lg\left(\frac{1}{5}\right) - \lg\left(\frac{1}{11}\right) = 1.138$
dept = low	$0 \cdot \lg\left(\frac{0}{4}\right) - \lg\left(\frac{1}{11}\right) = -\infty$
debt = high	$1 \cdot \lg\left(\frac{1}{7}\right) - \lg\left(\frac{1}{11}\right) = 0.652$
collateral = none	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{1}{11}\right) = 0.290$
collateral = adequate	$0 \cdot \lg\left(\frac{0}{2}\right) - \lg\left(\frac{1}{11}\right) = -\infty$
income = 0-15	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{1}{11}\right) = -\infty$

Candidate	Goodness
income = 15-35	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{1}{11}\right) = 1.874$
income = >35	$0 \cdot \lg\left(\frac{0}{5}\right) - \lg\left(\frac{1}{11}\right) = -\infty$

- IF income = 15-35 THEN risk = moderate $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{1}{3}\right) = 1.585]$

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

- IF income = 15-35 AND ? THEN risk = moderate $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{1}{3}\right) = 1.585]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$0 \cdot \lg\left(\frac{0}{7}\right) - \lg\left(\frac{1}{3}\right) = -\infty$
credit_history = unknown	$0 \cdot \lg\left(\frac{0}{7}\right) - \lg\left(\frac{1}{3}\right) = -\infty$
credit_history = good	$1 \cdot \lg\left(\frac{1}{7}\right) - \lg\left(\frac{1}{3}\right) = 1.585$
dept = low	$0 \cdot \lg\left(\frac{0}{9}\right) - \lg\left(\frac{1}{3}\right) = -\infty$
debt = high	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{1}{3}\right) = 0$
collateral = none	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{1}{3}\right) = 0$
collateral = adequate	$0 \cdot \lg\left(\frac{0}{9}\right) - \lg\left(\frac{1}{3}\right) = -\infty$

- IF income = 15-35 AND credit_history = good THEN risk = moderate [perfect]

This takes care of all the risk = moderate rules.

2.2.2.2.4 Resulting Rules

1. IF income = 15-35 AND debt = low THEN risk = moderate
2. IF credit_history = bad AND collateral = adequate THEN risk = moderate
3. IF income = 15-35 AND credit_history = good THEN risk = moderate

2.2.2.3 risk = high

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
1	bad	low	none	0-15	high
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
7	unknown	high	none	0-15	high
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low

Instance	credit_history	debt	collateral	income	risk
11	good	high	none	0-15	high
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.2.3.1 Rule 1

- IF ? THEN risk = high $[-\lg(\frac{P}{T}) = -\lg(\frac{5}{14}) = 1.485]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$2 \cdot \lg(\frac{2}{4}) - \lg(\frac{5}{14}) = 0.971$
credit_history = unknown	$5 \cdot \lg(\frac{2}{5}) - \lg(\frac{5}{14}) = 0.327$
credit_history = good	$1 \cdot \lg(\frac{1}{5}) - \lg(\frac{5}{14}) = -0.837$
dept = low	$1 \cdot \lg(\frac{1}{7}) - \lg(\frac{5}{14}) = -1.322$
debt = high	$4 \cdot \lg(\frac{4}{7}) - \lg(\frac{5}{14}) = 2.712$
collateral = none	$5 \cdot \lg(\frac{5}{11}) - \lg(\frac{5}{14}) = 1.740$
collateral = adequate	$0 \cdot \lg(\frac{0}{3}) - \lg(\frac{5}{14}) = -\infty$
income = 0-15	$3 \cdot \lg(\frac{3}{5}) - \lg(\frac{5}{14}) = 4.456$
income = 15-35	$2 \cdot \lg(\frac{2}{5}) - \lg(\frac{5}{14}) = 0.327$
income = >35	$0 \cdot \lg(\frac{0}{6}) - \lg(\frac{5}{14}) = -\infty$

- IF income = 0-15 THEN risk = high [perfect]

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

2.2.2.3.2 Rule 2

- IF ? THEN risk = high $[-\lg(\frac{P}{T}) = -\lg(\frac{2}{11}) = 2.459]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$1 \cdot \lg(\frac{1}{3}) - \lg(\frac{2}{11}) = 0.874$
credit_history = unknown	$1 \cdot \lg(\frac{1}{4}) - \lg(\frac{2}{11}) = 0.459$
credit_history = good	$0 \cdot \lg(\frac{0}{4}) - \lg(\frac{2}{11}) = -\infty$
dept = low	$0 \cdot \lg(\frac{0}{6}) - \lg(\frac{2}{11}) = -\infty$

Candidate	Goodness
debt = high	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{2}{11}\right) = 2.275$
collateral = none	$2 \cdot \lg\left(\frac{2}{8}\right) - \lg\left(\frac{2}{11}\right) = 0.919$
collateral = adequate	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{2}{11}\right) = -\infty$
income = 0-15	$0 \cdot \lg\left(\frac{0}{9}\right) - \lg\left(\frac{2}{11}\right) = -\infty$
income = 15-35	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{2}{11}\right) = 2.275$
income = >35	$0 \cdot \lg\left(\frac{0}{6}\right) - \lg\left(\frac{2}{11}\right) = -\infty$

- IF debt = high THEN risk = high $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{5}\right) = 1.322]$

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
10	good	high	adequate	>35	low
12	good	high	none	15-35	moderate
13	good	high	none	>35	low
14	bad	high	none	15-35	high

- IF debt = high AND ? THEN risk = high $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{5}\right) = 1.322]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$1 \cdot \lg\left(\frac{1}{7}\right) - \lg\left(\frac{2}{5}\right) = 1.322$
credit_history = unknown	$1 \cdot \lg\left(\frac{1}{7}\right) - \lg\left(\frac{2}{5}\right) = 1.322$
credit_history = good	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{2}{5}\right) = -\infty$
collateral = none	$2 \cdot \lg\left(\frac{2}{4}\right) - \lg\left(\frac{2}{5}\right) = 0.644$
collateral = adequate	$0 \cdot \lg\left(\frac{0}{7}\right) - \lg\left(\frac{2}{5}\right) = -\infty$
income = 0-15	$0 \cdot \lg\left(\frac{0}{9}\right) - \lg\left(\frac{2}{5}\right) = -\infty$
income = 15-35	$2 \cdot \lg\left(\frac{2}{5}\right) - \lg\left(\frac{2}{5}\right) = 1.474$
income = >35	$0 \cdot \lg\left(\frac{0}{2}\right) - \lg\left(\frac{2}{5}\right) = -\infty$

- IF debt = high AND income = 15-35 THEN risk = high $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{3}\right) = 0.585]$

Note: So far this is the only deviation from the rules produced when $\frac{P}{T}$ was used.

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
12	good	high	none	15-35	moderate
14	bad	high	none	15-35	high

- IF debt = high AND income = 15-35 AND ? THEN risk = high $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{2}{3}\right) = 0.585]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$1 \cdot \lg\left(\frac{1}{7}\right) - \lg\left(\frac{2}{3}\right) = 0.585$
credit_history = unknown	$1 \cdot \lg\left(\frac{1}{7}\right) - \lg\left(\frac{2}{3}\right) = 0.585$
credit_history = good	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{2}{3}\right) = -\infty$
collateral = none	$2 \cdot \lg\left(\frac{2}{3}\right) - \lg\left(\frac{2}{3}\right) = 0$

Candidate	Goodness
collateral = adequate	$0 \cdot \left \lg\left(\frac{0}{6}\right) - \lg\left(\frac{2}{3}\right) \right = -\infty$

- IF debt = high AND income = 15-35 AND credit_history = bad THEN risk = high [perfect]

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
4	bad	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low
8	bad	low	adequate	>35	moderate
9	good	low	none	>35	low
10	good	high	adequate	>35	low
12	good	high	none	15-35	moderate
13	good	high	none	>35	low

2.2.2.3.3 Rule 3

- IF ? THEN risk = high $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{1}{10}\right) = 3.322]$

Candidate Attributes	
Candidate	Goodness
credit_history = bad	$0 \cdot \left \lg\left(\frac{0}{2}\right) - \lg\left(\frac{1}{10}\right) \right = -\infty$
credit_history = unknown	$1 \cdot \left \lg\left(\frac{1}{4}\right) - \lg\left(\frac{1}{10}\right) \right = 1.322$
credit_history = good	$0 \cdot \left \lg\left(\frac{0}{4}\right) - \lg\left(\frac{1}{10}\right) \right = -\infty$
debt = low	$0 \cdot \left \lg\left(\frac{0}{6}\right) - \lg\left(\frac{1}{10}\right) \right = -\infty$
debt = high	$1 \cdot \left \lg\left(\frac{1}{4}\right) - \lg\left(\frac{1}{10}\right) \right = 1.322$
collateral = none	$1 \cdot \left \lg\left(\frac{1}{7}\right) - \lg\left(\frac{1}{10}\right) \right = 0.515$
collateral = adequate	$0 \cdot \left \lg\left(\frac{0}{3}\right) - \lg\left(\frac{1}{10}\right) \right = -\infty$
income = 0-15	$0 \cdot \left \lg\left(\frac{0}{0}\right) - \lg\left(\frac{1}{10}\right) \right = -\infty$
income = 15-35	$1 \cdot \left \lg\left(\frac{1}{4}\right) - \lg\left(\frac{1}{10}\right) \right = 1.322$
income = >35	$0 \cdot \left \lg\left(\frac{0}{6}\right) - \lg\left(\frac{1}{10}\right) \right = -\infty$

- IF credit_history = unknown THEN risk = high $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{1}{4}\right) = 2]$

Applicable Instances					
Instance	credit_history	debt	collateral	income	risk
2	unknown	high	none	15-35	high
3	unknown	low	none	15-35	moderate
5	unknown	low	adequate	>35	low
6	unknown	low	none	>35	low

- IF credit_history = unknown AND ? THEN risk = high $[-\lg\left(\frac{P}{T}\right) = -\lg\left(\frac{1}{4}\right) = 2]$

Candidate Attributes	
Candidate	Goodness
dept = low	$0 \cdot \lg\left(\frac{0}{3}\right) - \lg\left(\frac{1}{4}\right) = -\infty$
debt = high	$1 \cdot \lg\left(\frac{1}{1}\right) - \lg\left(\frac{1}{4}\right) = 2$
collateral = none	$1 \cdot \lg\left(\frac{1}{3}\right) - \lg\left(\frac{1}{4}\right) = 0.415$
collateral = adequate	$0 \cdot \lg\left(\frac{0}{1}\right) - \lg\left(\frac{1}{4}\right) = -\infty$
income = 0-15	$0 \cdot \lg\left(\frac{0}{0}\right) - \lg\left(\frac{1}{4}\right) = -\infty$
income = 15-35	$1 \cdot \lg\left(\frac{1}{2}\right) - \lg\left(\frac{1}{4}\right) = 1$
income = >35	$0 \cdot \lg\left(\frac{0}{2}\right) - \lg\left(\frac{1}{4}\right) = -\infty$

- IF credit_history = unknown AND debt = high THEN risk = high [perfect]

2.2.2.3.4 Resulting Rules

1. IF income = 0-15 THEN risk = high
2. IF debt = high AND income = 15-35 AND credit_history = bad THEN risk = high
3. IF credit_history = unknown AND debt = high THEN risk = high

2.2.2.4 Results

This concludes the computation of PRISM on this dataset using $p \cdot \left[\lg\left(\frac{p}{t}\right) - \lg\left(\frac{p}{T}\right)\right]$. There were 8 rules generated. They are:

1. risk = low
 - (a) IF income = >35 and credit_history = good THEN risk = low
 - (b) IF income = >35 AND credit_history = unknown THEN risk = low
2. risk = moderate
 - (a) IF income = 15-35 AND debt = low THEN risk = moderate
 - (b) IF credit_history = bad AND collateral = adequate THEN risk = moderate
 - (c) IF income = 15-35 AND credit_history = good THEN risk = moderate
3. risk = high
 - (a) IF income = 0-15 THEN risk = high
 - (b) IF debt = high AND income = 15-35 AND credit_history = bad THEN risk = high
 - (c) IF credit_history = unknown AND debt = high THEN risk = high

2.2.3 PRISM Pruning

Note: I can think of many ways in which to use this $m : RULES \rightarrow REALS$ function to prune rules. But because PRISM in WEKA doesn't seem to be doing any pruning, I am assuming that the project description means to describe or explain the pruning method shown in the class text on pages 175-177.

Firstly the PRISM algorithm is used to generate a perfect rule for the first target attribute value. This rule is put through m in order to compute the likelihood of the rule being as good as it is purely by chance. Then the last attribute=value pair test is removed from that rule and the m function is computed over this shortened rule once more. Now the two computed values are compared. If the longer rule is less likely to occur by chance then it is kept and PRISM does its work again to get another perfect rule.

If, however, it is the case that the longer rule has a larger likelihood to occur by chance than the shorted rule then there is no purpose in keeping the longer rule. The short one is deemed more rewarding as it is

less likely to classify correctly by chance. The shortening process is then repeated again on the shorter rule until either shortening doesn't produce a better (less likely to occur by chance) rule or there is nothing left to remove. After the rule is optimized, PRISM begins its work again to come up with another perfect rule as if it generated the optimized rule all by its own.

Another similar method might be for PRISM to compare the chance values of a rule it is working on currently and the rule with an attribute added that it is considering to add. This way it could stop adding tests seems to make the rule more likely to occur by chance. The same $m : RULES \rightarrow REALS$ could be used. The two possible methods might not be similar. If one looks at a perfect rule and all possible "prefix" rules of that rule, there might be hills in the chance value these rules receive from the function. That is going from a full rule to a shorter rule might make the prune stop at one point, but going from an empty rule to a longer rule might make the rule creation stop at a different point.