

# Project 3: Data Pre-processing, Mining, and Evaluation of Association Rules: Homework

Piotr Mardziel

14th April 2004

# Contents

<b>3</b>	<b>Project 3: Data Pre-processing, Mining, and Evaluation of Association Rules: Homework</b>	<b>2</b>
3.1	Group Members . . . . .	2
3.2	Pre-Project Homework Assignment . . . . .	2
3.2.1	Dataset . . . . .	2
3.2.2	Candidate Generation . . . . .	3
3.2.2.1	Size 1 . . . . .	3
3.2.2.2	Size 2 . . . . .	4
3.2.2.3	Size 3 . . . . .	9
3.2.2.4	Size 4 . . . . .	11
3.2.2.5	The Candidates . . . . .	11
3.2.3	Association Rule Generation . . . . .	12
3.2.3.1	2-Element Frequent Sets . . . . .	12
3.2.3.2	Rules So Far . . . . .	17
3.2.3.3	3-Element Frequent Sets . . . . .	17
3.2.3.4	Totals . . . . .	21

## Chapter 3

# Project 3: Data Pre-processing, Mining, and Evaluation of Association Rules: Homework

### 3.1 Group Members

GROUP MEMBER	USERNAME	TASKS
Piotr Mardziel	piotrm	everything

### 3.2 Pre-Project Homework Assignment

#### 3.2.1 Dataset

The dataset is comprised of some credit information. All values were marked with their attribute names beforehand. It is also noted that a fraction of at least  $\frac{3}{14}$  is required in the case of this dataset to achieve the minimum support. This translates to 3 instances minimum for any candidate set.

- {credit\_history=bad, debt=low, collateral=none, income=0-15, risk=high}
- {credit\_history=unknown, debt=high, collateral=none, income=15-35, risk=high}
- {credit\_history=unknown, debt=low, collateral=none, income=15-35, risk=moderate}
- {credit\_history=bad, debt=low, collateral=none, income=15-35, risk=moderate}
- {credit\_history=unknown, debt=low, collateral=adequate, income=>35, risk=low}
- {credit\_history=unknown, debt=low, collateral=none, income=>35, risk=low}
- {credit\_history=unknown, debt=high, collateral=none, income=0-15, risk=high}
- {credit\_history=bad, debt=low, collateral=adequate, income=>35, risk=moderate}
- {credit\_history=good, debt=low, collateral=none, income=>35, risk=low}
- {credit\_history=good, debt=high, collateral=adequate, income=>35, risk=low}

- {credit\_history=good, debt=high, collateral=none, income=0-15, risk=high}
- {credit\_history=good, debt=high, collateral=none, income=15-35, risk=moderate}
- {credit\_history=good, debt=high, collateral=none, income=>35, risk=low}
- {credit\_history=bad, debt=high, collateral=none, income=15-35, risk=high}

### 3.2.2 Candidate Generation

#### 3.2.2.1 Size 1

3.2.2.1.1  $C_1$  Initially all possible attribute and their possible values are candidates.

- {credit\_history=bad}
- {credit\_history=unknown}
- {credit\_history=good}
- {debt=low}
- {debt=high}
- {collateral=none}
- {collateral=adequate}
- {income=0-15}
- {income=15-35}
- {income=>35}
- {risk=low}
- {risk=moderate}
- {risk=high}

3.2.2.1.2  $F_1$  Now the actual supports are calculated from counting the matched instances in the dataset. Note that boldface denotes satisfactory support.

CANDIDATE	SUPPORT
{credit_history=bad}	4/14
{credit_history=unknown}	5/14
{credit_history=good}	5/14
{debt=low}	7/14
{debt=high}	7/14
{collateral=none}	11/14
{collateral=adequate}	3/14
{income=0-15}	3/14
{income=15-35}	5/14
{income=>35}	6/14
{risk=low}	5/14
{risk=moderate}	4/14
{risk=high}	5/14

Looks like everyone stays. Note that at least 3 instances are required to muster enough support. Everyone passes.

### 3.2.2.2 Size 2

**3.2.2.2.1**  $C_2$  Now each combination of two attributes is a possible candidate and since none got eliminated in the previous round, there are a lot of candidates now. Note that although it is silly to assume the existence of instances with two different `credit_history` values in this case, the algorithm has no knowledge of the relationship between `credit_history=bad` and `credit_history=unknown` or others in the same form. As far as it is concerned, these are just different elements that can or can not be in a candidate set.

- {`credit_history=bad`, `credit_history=unknown`}
- {`credit_history=bad`, `credit_history=good`}
- {`credit_history=bad`, `debt=low`}
- {`credit_history=bad`, `debt=high`}
- {`credit_history=bad`, `collateral=none`}
- {`credit_history=bad`, `collateral=adequate`}
- {`credit_history=bad`, `income=0-15`}
- {`credit_history=bad`, `income=15-35`}
- {`credit_history=bad`, `income=>35`}
- {`credit_history=bad`, `risk=low`}
- {`credit_history=bad`, `risk=moderate`}
- {`credit_history=bad`, `risk=high`}
- {`credit_history=unknown`, `credit_history=good`}
- {`credit_history=unknown`, `debt=low`}
- {`credit_history=unknown`, `debt=high`}
- {`credit_history=unknown`, `collateral=none`}
- {`credit_history=unknown`, `collateral=adequate`}
- {`credit_history=unknown`, `income=0-15`}
- {`credit_history=unknown`, `income=15-35`}
- {`credit_history=unknown`, `income=>35`}
- {`credit_history=unknown`, `risk=low`}
- {`credit_history=unknown`, `risk=moderate`}
- {`credit_history=unknown`, `risk=high`}
- {`credit_history=good`, `debt=low`}
- {`credit_history=good`, `debt=high`}

- {credit\_history=good, collateral=none}
- {credit\_history=good, collateral=adequate}
- {credit\_history=good, income=0-15}
- {credit\_history=good, income=15-35}
- {credit\_history=good, income=>35}
- {credit\_history=good, risk=low}
- {credit\_history=good, risk=moderate}
- {credit\_history=good, risk=high}
- {debt=low, debt=high}
- {debt=low, collateral=none}
- {debt=low, collateral=adequate}
- {debt=low, income=0-15}
- {debt=low, income=15-35}
- {debt=low, income=>35}
- {debt=low, risk=low}
- {debt=low, risk=moderate}
- {debt=low, risk=high}
- {debt=high, collateral=none}
- {debt=high, collateral=adequate}
- {debt=high, income=0-15}
- {debt=high, income=15-35}
- {debt=high, income=>35}
- {debt=high, risk=low}
- {debt=high, risk=moderate}
- {debt=high, risk=high}
- {collateral=none, collateral=adequate}
- {collateral=none, income=0-15}
- {collateral=none, income=15-35}
- {collateral=none, income=>35}
- {collateral=none, risk=low}
- {collateral=none, risk=moderate}

- {collateral=none, risk=high}
- {collateral=adequate, income=0-15}
- {collateral=adequate, income=15-35}
- {collateral=adequate, income=>35}
- {collateral=adequate, risk=low}
- {collateral=adequate, risk=moderate}
- {collateral=adequate, risk=high}
- {income=0-15, income=15-35}
- {income=0-15, income=>35}
- {income=0-15, risk=low}
- {income=0-15, risk=moderate}
- {income=0-15, risk=high}
- {income=15-35, income=>35}
- {income=15-35, risk=low}
- {income=15-35, risk=moderate}
- {income=15-35, risk=high}
- {income=>35, risk=low}
- {income=>35, risk=moderate}
- {income=>35, risk=high}
- {risk=low, risk=moderate}
- {risk=low, risk=high}
- {risk=moderate, risk=high}

**3.2.2.2.2**  $F_2$  In order to prune a lot of these, their support is computed.

CANDIDATE	SUPPORT
{credit_history=bad, credit_history=unknown}	0/14
{credit_history=bad, credit_history=good}	0/14
<b>{credit_history=bad, debt=low}</b>	<b>3/14</b>
{credit_history=bad, debt=high}	1/14
<b>{credit_history=bad, collateral=none}</b>	<b>3/14</b>
{credit_history=bad, collateral=adequate}	1/14
{credit_history=bad, income=0-15}	1/14
{credit_history=bad, income=15-35}	2/14
{credit_history=bad, income=>35}	1/14
{credit_history=bad, risk=low}	0/14
{credit_history=bad, risk=moderate}	2/14

CANDIDATE	SUPPORT
{credit_history=bad, risk=high}	2/14
{credit_history=unknown, credit_history=good}	0/14
<b>{credit_history=unknown, debt=low}</b>	<b>3/14</b>
{credit_history=unknown, debt=high}	2/14
<b>{credit_history=unknown, collateral=none}</b>	<b>4/14</b>
{credit_history=unknown, collateral=adequate}	1/14
{credit_history=unknown, income=0-15}	1/14
{credit_history=unknown, income=15-35}	2/14
{credit_history=unknown, income=>35}	2/14
{credit_history=unknown, risk=low}	2/14
{credit_history=unknown, risk=moderate}	1/14
{credit_history=unknown, risk=high}	2/14
{credit_history=good, debt=low}	1/14
<b>{credit_history=good, debt=high}</b>	<b>4/14</b>
<b>{credit_history=good, collateral=none}</b>	<b>4/14</b>
{credit_history=good, collateral=adequate}	1/14
{credit_history=good, income=0-15}	1/14
{credit_history=good, income=15-35}	1/14
<b>{credit_history=good, income=&gt;35}</b>	<b>3/14</b>
<b>{credit_history=good, risk=low}</b>	<b>3/14</b>
{credit_history=good, risk=moderate}	1/14
{credit_history=good, risk=high}	1/14
{debt=low, debt=high}	0/14
<b>{debt=low, collateral=none}</b>	<b>5/14</b>
{debt=low, collateral=adequate}	2/14
{debt=low, income=0-15}	1/14
{debt=low, income=15-35}	2/14
<b>{debt=low, income=&gt;35}</b>	<b>4/14</b>
<b>{debt=low, risk=low}</b>	<b>3/14</b>
<b>{debt=low, risk=moderate}</b>	<b>3/14</b>
{debt=low, risk=high}	1/14
<b>{debt=high, collateral=none}</b>	<b>6/14</b>
{debt=high, collateral=adequate}	1/14
{debt=high, income=0-15}	2/14
<b>{debt=high, income=15-35}</b>	<b>3/14</b>
{debt=high, income=>35}	2/14
{debt=high, risk=low}	2/14
{debt=high, risk=moderate}	1/14
<b>{debt=high, risk=high}</b>	<b>4/14</b>
{collateral=none, collateral=adequate}	0/14
<b>{collateral=none, income=0-15}</b>	<b>3/14</b>
<b>{collateral=none, income=15-35}</b>	<b>5/14</b>
<b>{collateral=none, income=&gt;35}</b>	<b>3/14</b>
<b>{collateral=none, risk=low}</b>	<b>3/14</b>
<b>{collateral=none, risk=moderate}</b>	<b>3/14</b>
<b>{collateral=none, risk=high}</b>	<b>5/14</b>
{collateral=adequate, income=0-15}	0/14
{collateral=adequate, income=15-35}	0/14



CANDIDATE	SUPPORT
<b>{collateral=adequate, income=&gt;35}</b>	<b>3/14</b>
{collateral=adequate, risk=low}	2/14
{collateral=adequate, risk=moderate}	1/14
{collateral=adequate, risk=high}	0/14
{income=0-15, income=15-35}	0/14
{income=0-15, income=>35}	0/14
{income=0-15, risk=low}	0/14
{income=0-15, risk=moderate}	0/14
<b>{income=0-15, risk=high}</b>	<b>3/14</b>
{income=15-35, income=>35}	0/14
{income=15-35, risk=low}	0/14
<b>{income=15-35, risk=moderate}</b>	<b>3/14</b>
{income=15-35, risk=high}	2/14
<b>{income=&gt;35, risk=low}</b>	<b>5/14</b>
{income=>35, risk=moderate}	1/14
{income=>35, risk=high}	0/14
{risk=low, risk=moderate}	0/14
{risk=low, risk=high}	0/14
{risk=moderate, risk=high}	0/14

After removing the sets that don't have enough support, the only things left are:

CANDIDATE	SUPPORT
<b>{credit_history=bad, debt=low}</b>	<b>3/14</b>
<b>{credit_history=bad, collateral=none}</b>	<b>3/14</b>
<b>{credit_history=unknown, debt=low}</b>	<b>3/14</b>
<b>{credit_history=unknown, collateral=none}</b>	<b>4/14</b>
<b>{credit_history=good, debt=high}</b>	<b>4/14</b>
<b>{credit_history=good, collateral=none}</b>	<b>4/14</b>
<b>{credit_history=good, income=&gt;35}</b>	<b>3/14</b>
<b>{credit_history=good, risk=low}</b>	<b>3/14</b>
<b>{debt=low, collateral=none}</b>	<b>5/14</b>
<b>{debt=low, income=&gt;35}</b>	<b>4/14</b>
<b>{debt=low, risk=low}</b>	<b>3/14</b>
<b>{debt=low, risk=moderate}</b>	<b>3/14</b>
<b>{debt=high, collateral=none}</b>	<b>6/14</b>
<b>{debt=high, income=15-35}</b>	<b>3/14</b>
<b>{debt=high, risk=high}</b>	<b>4/14</b>
<b>{collateral=none, income=0-15}</b>	<b>3/14</b>
<b>{collateral=none, income=15-35}</b>	<b>5/14</b>
<b>{collateral=none, income=&gt;35}</b>	<b>3/14</b>
<b>{collateral=none, risk=low}</b>	<b>3/14</b>
<b>{collateral=none, risk=moderate}</b>	<b>3/14</b>
<b>{collateral=none, risk=high}</b>	<b>5/14</b>
<b>{collateral=adequate, income=&gt;35}</b>	<b>3/14</b>
<b>{income=0-15, risk=high}</b>	<b>3/14</b>
<b>{income=15-35, risk=moderate}</b>	<b>3/14</b>

CANDIDATE	SUPPORT
{income=>35, risk=low}	5/14

### 3.2.2.3 Size 3

**3.2.2.3.1**  $C_3$  Now to produce candidates of 3 elements, one considers all possible pairs sets from  $F_2$  that differ only in their last elements. These are merged together and the result is:

- {credit\_history=bad, debt=low, collateral=none}
- {credit\_history=unknown, debt=low, collateral=none}
- {credit\_history=good, debt=high, collateral=none}
- \* {credit\_history=good, debt=high, income=>35}
- \* {credit\_history=good, debt=high, risk=low}
- {credit\_history=good, collateral=none, income=>35}
- {credit\_history=good, collateral=none, risk=low}
- {credit\_history=good, income=>35, risk=low}
- {debt=low, collateral=none, income=>35}
- {debt=low, collateral=none, risk=low}
- {debt=low, collateral=none, risk=moderate}
- {debt=low, income=>35, risk=low}
- \* {debt=low, income=>35, risk=moderate}
- \* {debt=low, risk=low, risk=moderate}
- {debt=high, collateral=none, income=15-35}
- {debt=high, collateral=none, risk=high}
- \* {debt=high, income=15-35, risk=high}
- \* {collateral=none, income=0-15, income=15-35}
- \* {collateral=none, income=0-15, income=>35}
- \* {collateral=none, income=0-15, risk=low}
- \* {collateral=none, income=0-15, risk=moderate}
- {collateral=none, income=0-15, risk=high}
- \* {collateral=none, income=15-35, income=>35}
- \* {collateral=none, income=15-35, risk=low}
- {collateral=none, income=15-35, risk=moderate}
- \* {collateral=none, income=15-35, risk=high}

- {collateral=none, income=>35, risk=low}
- \* {collateral=none, income=>35, risk=moderate}
- \* {collateral=none, income=>35, risk=high}
- \* {collateral=none, risk=low, risk=moderate}
- \* {collateral=none, risk=low, risk=high}
- \* {collateral=none, risk=moderate, risk=high}

The ones marked with \* are not considered because their one one their subsets of 2 elements (mainly the last two elements) do not belong to  $F_2$ . So in reality only the following are candidates:

- {credit\_history=bad, debt=low, collateral=none}
- {credit\_history=unknown, debt=low, collateral=none}
- {credit\_history=good, debt=high, collateral=none}
- {credit\_history=good, collateral=none, income=>35}
- {credit\_history=good, collateral=none, risk=low}
- {credit\_history=good, income=>35, risk=low}
- {debt=low, collateral=none, income=>35}
- {debt=low, collateral=none, risk=low}
- {debt=low, collateral=none, risk=moderate}
- {debt=low, income=>35, risk=low}
- {debt=high, collateral=none, income=15-35}
- {debt=high, collateral=none, risk=high}
- {collateral=none, income=0-15, risk=high}
- {collateral=none, income=15-35, risk=moderate}
- {collateral=none, income=>35, risk=low}

**3.2.2.3.2**  $F_3$  Now once more the support of all of these needs to be computed.

CANDIDATE	SUPPORT
{credit_history=bad, debt=low, collateral=none}	2/14
{credit_history=unknown, debt=low, collateral=none}	2/14
<b>{credit_history=good, debt=high, collateral=none}</b>	<b>3/14</b>
{credit_history=good, collateral=none, income=>35}	2/14
{credit_history=good, collateral=none, risk=low}	2/14
<b>{credit_history=good, income=&gt;35, risk=low}</b>	<b>3/14</b>
{debt=low, collateral=none, income=>35}	2/14
{debt=low, collateral=none, risk=low}	2/14
{debt=low, collateral=none, risk=moderate}	2/14

CANDIDATE	SUPPORT
{debt=low, income=>35, risk=low}	3/14
{debt=high, collateral=none, income=15-35}	3/14
{debt=high, collateral=none, risk=high}	4/14
{collateral=none, income=0-15, risk=high}	3/14
{collateral=none, income=15-35, risk=moderate}	3/14
{collateral=none, income=>35, risk=low}	3/14

Removing the guys that didn't make it gives the following table.

CANDIDATE	SUPPORT
{credit_history=good, debt=high, collateral=none}	3/14
{credit_history=good, income=>35, risk=low}	3/14
{debt=low, income=>35, risk=low}	3/14
{debt=high, collateral=none, income=15-35}	3/14
{debt=high, collateral=none, risk=high}	4/14
{collateral=none, income=0-15, risk=high}	3/14
{collateral=none, income=15-35, risk=moderate}	3/14
{collateral=none, income=>35, risk=low}	3/14

### 3.2.2.4 Size 4

**3.2.2.4.1**  $C_4$  Now the elements from  $F_3$  that differ only in their last element are merged. The result is a rather small candidate set.

{debt=high, collateral=none, income=15-35, risk=high}

Furthermore, this isn't even considered by the algorithm since {collateral=none, income=15-35, risk=high} is not an element of  $F_3$ .

**3.2.2.4.2**  $F_4$  Nothing is even considered for  $F_4$  so we stop here with candidate generation.

### 3.2.2.5 The Candidates

Here they are, the candidates of all the sizes.

CANDIDATE	SUPPORT
{credit_history=bad}	4/14
{credit_history=unknown}	5/14
{credit_history=good}	5/14
{debt=low}	7/14
{debt=high}	7/14
{collateral=none}	11/14
{collateral=adequate}	3/14
{income=0-15}	3/14
{income=15-35}	5/14
{income=>35}	6/14
{risk=low}	5/14
{risk=moderate}	4/14
{risk=high}	5/14

CANDIDATE	SUPPORT
{credit_history=bad, debt=low}	3/14
{credit_history=bad, collateral=none}	3/14
{credit_history=unknown, debt=low}	3/14
{credit_history=unknown, collateral=none}	4/14
{credit_history=good, debt=high}	4/14
{credit_history=good, collateral=none}	4/14
{credit_history=good, income=>35}	3/14
{credit_history=good, risk=low}	3/14
{debt=low, collateral=none}	5/14
{debt=low, income=>35}	4/14
{debt=low, risk=low}	3/14
{debt=low, risk=moderate}	3/14
{debt=high, collateral=none}	6/14
{debt=high, income=15-35}	3/14
{debt=high, risk=high}	4/14
{collateral=none, income=0-15}	3/14
{collateral=none, income=15-35}	5/14
{collateral=none, income=>35}	3/14
{collateral=none, risk=low}	3/14
{collateral=none, risk=moderate}	3/14
{collateral=none, risk=high}	5/14
{collateral=adequate, income=>35}	3/14
{income=0-15, risk=high}	3/14
{income=15-35, risk=moderate}	3/14
{income=>35, risk=low}	5/14
{credit_history=good, debt=high, collateral=none}	3/14
{credit_history=good, income=>35, risk=low}	3/14
{debt=low, income=>35, risk=low}	3/14
{debt=high, collateral=none, income=15-35}	3/14
{debt=high, collateral=none, risk=high}	4/14
{collateral=none, income=0-15, risk=high}	3/14
{collateral=none, income=15-35, risk=moderate}	3/14
{collateral=none, income=>35, risk=low}	3/14

### 3.2.3 Association Rule Generation

Now all the candidates are present and all the possible useful supports computed, it is time to actually look for rules that are precise enough.

The algorithm starts us first with a loop across all frequent sets of 2 elements and forms an association rule from taking one of the elements out in order to make up the consequent. The rest becomes the antecedent.

#### 3.2.3.1 2-Element Frequent Sets

- {credit\_history=bad, debt=low}
  - {debt=low}  $\implies$  {credit\_history=bad}
    - \*  $\frac{s(\{credit\_history=bad,debt=low\})}{s(\{credit\_history=bad\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
  - {credit\_history=bad}  $\implies$  {debt=low}

- \*  $\frac{s(\{\text{credit\_history}=\text{bad}, \text{debt}=\text{low}\})}{s(\{\text{debt}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
- {credit\_history=bad, collateral=none}
  - {collateral=none}  $\implies$  {credit\_history=bad}
    - \*  $\frac{s(\{\text{credit\_history}=\text{bad}, \text{collateral}=\text{none}\})}{s(\{\text{credit\_history}=\text{bad}\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
  - {credit\_history=bad}  $\implies$  {collateral=none}
    - \*  $\frac{s(\{\text{credit\_history}=\text{bad}, \text{collateral}=\text{none}\})}{s(\{\text{collateral}=\text{none}\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9$
- {credit\_history=unknown, debt=low}
  - {debt=low}  $\implies$  {credit\_history=unknown}
    - \*  $\frac{s(\{\text{credit\_history}=\text{unknown}, \text{debt}=\text{low}\})}{s(\{\text{credit\_history}=\text{unknown}\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
  - {credit\_history=unknown}  $\implies$  {debt=low}
    - \*  $\frac{s(\{\text{credit\_history}=\text{unknown}, \text{debt}=\text{low}\})}{s(\{\text{debt}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
- {credit\_history=unknown, collateral=none}
  - {collateral=none}  $\implies$  {credit\_history=unknown}
    - \*  $\frac{s(\{\text{credit\_history}=\text{unknown}, \text{collateral}=\text{none}\})}{s(\{\text{collateral}=\text{none}\})} = \frac{\frac{4}{14}}{\frac{11}{14}} = \frac{4}{11} < 0.9$
  - {credit\_history=unknown}  $\implies$  {collateral=none}
    - \*  $\frac{s(\{\text{credit\_history}=\text{unknown}, \text{collateral}=\text{none}\})}{s(\{\text{credit\_history}=\text{unknown}\})} = \frac{\frac{4}{14}}{\frac{5}{14}} = \frac{4}{5} < 0.9$
- {credit\_history=good, debt=high}
  - {debt=high}  $\implies$  {credit\_history=good}
    - \*  $\frac{s(\{\text{credit\_history}=\text{good}, \text{debt}=\text{high}\})}{s(\{\text{debt}=\text{high}\})} = \frac{\frac{4}{14}}{\frac{7}{14}} = \frac{4}{7} < 0.9$
  - {credit\_history=good}  $\implies$  {debt=high}
    - \*  $\frac{s(\{\text{credit\_history}=\text{good}, \text{debt}=\text{high}\})}{s(\{\text{credit\_history}=\text{good}\})} = \frac{\frac{4}{14}}{\frac{5}{14}} = \frac{4}{5} < 0.9$
- {credit\_history=good, collateral=none}
  - {collateral=none}  $\implies$  {credit\_history=good}
    - \*  $\frac{s(\{\text{credit\_history}=\text{good}, \text{collateral}=\text{none}\})}{s(\{\text{collateral}=\text{none}\})} = \frac{\frac{4}{14}}{\frac{11}{14}} = \frac{4}{11} < 0.9$
  - {credit\_history=good}  $\implies$  {collateral=none}
    - \*  $\frac{s(\{\text{credit\_history}=\text{good}, \text{collateral}=\text{none}\})}{s(\{\text{credit\_history}=\text{good}\})} = \frac{\frac{4}{14}}{\frac{5}{14}} = \frac{4}{5} < 0.9$
- {credit\_history=good, income=>35}
  - {income=>35}  $\implies$  {credit\_history=good}
    - \*  $\frac{s(\{\text{credit\_history}=\text{good}, \text{income}=>35\})}{s(\{\text{income}=>35\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$

- {credit\_history=good}  $\implies$  {income=>35}
  - \*  $\frac{s(\{credit\_history=good, income=>35\})}{s(\{credit\_history=good\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
- {credit\_history=good, risk=low}
  - {risk=low}  $\implies$  {credit\_history=good}
    - \*  $\frac{s(\{credit\_history=good, risk=low\})}{s(\{risk=low\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
  - {credit\_history=good}  $\implies$  {risk=low}
    - \*  $\frac{s(\{credit\_history=good, risk=low\})}{s(\{credit\_history=good\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
- {debt=low, collateral=none}
  - {collateral=none}  $\implies$  {debt=low}
    - \*  $\frac{s(\{debt=low, collateral=none\})}{s(\{collateral=none\})} = \frac{\frac{5}{14}}{\frac{11}{14}} = \frac{5}{11} < 0.9$
  - {debt=low}  $\implies$  {collateral=none}
    - \*  $\frac{s(\{debt=low, collateral=none\})}{s(\{debt=low\})} = \frac{\frac{5}{14}}{\frac{7}{14}} = \frac{5}{7} < 0.9$
- {debt=low, income=>35}
  - {income=>35}  $\implies$  {debt=low}
    - \*  $\frac{s(\{debt=low, income=>35\})}{s(\{income=>35\})} = \frac{\frac{4}{14}}{\frac{6}{14}} = \frac{4}{6} < 0.9$
  - {debt=low}  $\implies$  {income=>35}
    - \*  $\frac{s(\{debt=low, income=>35\})}{s(\{debt=low\})} = \frac{\frac{4}{14}}{\frac{7}{14}} = \frac{4}{7} < 0.9$
- {debt=low, risk=low}
  - {risk=low}  $\implies$  {debt=low}
    - \*  $\frac{s(\{debt=low, risk=low\})}{s(\{risk=low\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
  - {debt=low}  $\implies$  {risk=low}
    - \*  $\frac{s(\{debt=low, risk=low\})}{s(\{debt=low\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
- {debt=low, risk=moderate}
  - {risk=moderate}  $\implies$  {debt=low}
    - \*  $\frac{s(\{debt=low, risk=moderate\})}{s(\{risk=moderate\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
  - {debt=low}  $\implies$  {risk=moderate}
    - \*  $\frac{s(\{debt=low, risk=moderate\})}{s(\{debt=low\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
- {debt=high, collateral=none}
  - {collateral=none}  $\implies$  {debt=high}

- \*  $\frac{s(\{debt=high, collateral=none\})}{s(\{collateral=none\})} = \frac{\frac{6}{14}}{\frac{11}{14}} = \frac{6}{11} < 0.9$
- $\{debt=high\} \implies \{collateral=none\}$
- \*  $\frac{s(\{debt=high, collateral=none\})}{s(\{debt=high\})} = \frac{\frac{6}{14}}{\frac{7}{14}} = \frac{6}{7} < 0.9$
- $\{debt=high, income=15-35\}$ 
  - $\{income=15-35\} \implies \{debt=high\}$
  - \*  $\frac{s(\{debt=high, income=15-35\})}{s(\{income=15-35\})} = \frac{\frac{3}{14}}{\frac{9}{14}} = \frac{3}{9} < 0.9$
  - $\{debt=high\} \implies \{income=15-35\}$
  - \*  $\frac{s(\{debt=high, income=15-35\})}{s(\{debt=high\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
- $\{debt=high, risk=high\}$ 
  - $\{risk=high\} \implies \{debt=high\}$
  - \*  $\frac{s(\{debt=high, risk=high\})}{s(\{risk=high\})} = \frac{\frac{4}{14}}{\frac{5}{14}} = \frac{4}{5} < 0.9$
  - $\{debt=high\} \implies \{risk=high\}$
  - \*  $\frac{s(\{debt=high, risk=high\})}{s(\{debt=high\})} = \frac{\frac{4}{14}}{\frac{7}{14}} = \frac{4}{7} < 0.9$
- $\{collateral=none, income=0-15\}$ 
  - $\{income=0-15\} \implies \{collateral=none\}$
  - \*  $\frac{s(\{collateral=none, income=0-15\})}{s(\{income=0-15\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
  - $\{collateral=none\} \implies \{income=0-15\}$
  - \*  $\frac{s(\{collateral=none, income=0-15\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9$
- $\{collateral=none, income=15-35\}$ 
  - $\{income=15-35\} \implies \{collateral=none\}$
  - \*  $\frac{s(\{collateral=none, income=15-35\})}{s(\{income=15-35\})} = \frac{\frac{5}{14}}{\frac{5}{14}} = 1 \geq 0.9$
  - $\{collateral=none\} \implies \{income=15-35\}$
  - \*  $\frac{s(\{collateral=none, income=15-35\})}{s(\{collateral=none\})} = \frac{\frac{5}{14}}{\frac{11}{14}} = \frac{5}{11} < 0.9$
- $\{collateral=none, income=>35\}$ 
  - $\{income=>35\} \implies \{collateral=none\}$
  - \*  $\frac{s(\{collateral=none, income=>35\})}{s(\{income=>35\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
  - $\{collateral=none\} \implies \{income=>35\}$
  - \*  $\frac{s(\{collateral=none, income=>35\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9$
- $\{collateral=none, risk=low\}$



- {risk=low}  $\implies$  {collateral=none}
  - \*  $\frac{s(\{collateral=none,risk=low\})}{s(\{risk=low\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
- {collateral=none}  $\implies$  {risk=low}
  - \*  $\frac{s(\{collateral=none,risk=low\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9$
- {collateral=none, risk=moderate}
  - {risk=moderate}  $\implies$  {collateral=none}
    - \*  $\frac{s(\{collateral=none,risk=moderate\})}{s(\{risk=moderate\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
  - {collateral=none}  $\implies$  {risk=moderate}
    - \*  $\frac{s(\{collateral=none,risk=moderate\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9$
- {collateral=none, risk=high}
  - {risk=high}  $\implies$  {collateral=none}
    - \*  $\frac{s(\{collateral=none,risk=high\})}{s(\{risk=high\})} = \frac{\frac{5}{14}}{\frac{5}{14}} = 1 \geq 0.9$
  - {collateral=none}  $\implies$  {risk=high}
    - \*  $\frac{s(\{collateral=none,risk=high\})}{s(\{collateral=none\})} = \frac{\frac{5}{14}}{\frac{11}{14}} = \frac{5}{11} < 0.9$
- {collateral=adequate, income=>35}
  - {income=>35}  $\implies$  {collateral=adequate}
    - \*  $\frac{s(\{collateral=adequate,income=>35\})}{s(\{income=>35\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
  - {collateral=adequate}  $\implies$  {income=>35}
    - \*  $\frac{s(\{collateral=adequate,income=>35\})}{s(\{collateral=adequate\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
- {income=0-15, risk=high}
  - {risk=high}  $\implies$  {income=0-15}
    - \*  $\frac{s(\{income=0-15,risk=high\})}{s(\{risk=high\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
  - {income=0-15}  $\implies$  {risk=high}
    - \*  $\frac{s(\{income=0-15,risk=high\})}{s(\{income=0-15\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
- {income=15-35, risk=moderate}
  - {risk=moderate}  $\implies$  {income=15-35}
    - \*  $\frac{s(\{income=15-35,risk=moderate\})}{s(\{risk=moderate\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
  - {income=15-35}  $\implies$  {risk=moderate}
    - \*  $\frac{s(\{income=15-35,risk=moderate\})}{s(\{income=15-35\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
- {income=>35, risk=low}

- {risk=low}  $\implies$  {income=>35}
  - \*  $\frac{s(\{income=>35,risk=low\})}{s(\{risk=low\})} = \frac{\frac{5}{14}}{\frac{5}{14}} = 1 \geq 0.9$
- {income=>35}  $\implies$  {risk=low}
  - \*  $\frac{s(\{income=>35,risk=low\})}{s(\{income=>35\})} = \frac{\frac{5}{14}}{\frac{6}{14}} = \frac{5}{6} < 0.9$

### 3.2.3.2 Rules So Far

The frequent sets of 2 elements have so far generated the following six association rules.

1. {income=0-15}  $\implies$  {collateral=none}
2. {income=15-35}  $\implies$  {collateral=none}
3. {risk=high}  $\implies$  {collateral=none}
4. {collateral=adequate}  $\implies$  {income=>35}
5. {income=0-15}  $\implies$  {risk=high}
6. {risk=low}  $\implies$  {income=>35}

### 3.2.3.3 3-Element Frequent Sets

Now it is time to consider the 3 element frequent sets and the rules they can produce by splitting into antecedents and consequents.

- {credit\_history=good, debt=high, collateral=none}
  - 1 consequent
    - \* {debt=high, collateral=none}  $\implies$  {credit\_history=good}
      - $\frac{s(\{credit\_history=good,debt=high,collateral=none\})}{s(\{debt=high,collateral=none\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
    - \* {credit\_history=good, collateral=none}  $\implies$  {debt=high}
      - $\frac{s(\{credit\_history=good,debt=high,collateral=none\})}{s(\{credit\_history=good,collateral=none\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
    - \* {credit\_history=good, debt=high}  $\implies$  {collateral=none}
      - $\frac{s(\{credit\_history=good,debt=high,collateral=none\})}{s(\{credit\_history=good,debt=high\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
  - 2 consequents
    - \* {collateral=none}  $\implies$  {credit\_history=good, debt=high}
      - $\frac{s(\{credit\_history=good,debt=high,collateral=none\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9$
    - \* {debt=high}  $\implies$  {credit\_history=good, collateral=none}
      - $\frac{s(\{credit\_history=good,debt=high,collateral=none\})}{s(\{debt=high\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
    - \* {credit\_history=good}  $\implies$  {debt=high, collateral=none}
      - $\frac{s(\{credit\_history=good,debt=high,collateral=none\})}{s(\{credit\_history=good\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
- {credit\_history=good, income=>35, risk=low}
  - 1 consequent

- \*  $\{\text{income}=>35, \text{risk}=\text{low}\} \implies \{\text{credit\_history}=\text{good}\}$ 
  - $\frac{s(\{\text{credit\_history}=\text{good}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{income}=>35, \text{risk}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
- \*  $\{\text{credit\_history}=\text{good}, \text{risk}=\text{low}\} \implies \{\text{income}=>35\}$ 
  - $\frac{s(\{\text{credit\_history}=\text{good}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{credit\_history}=\text{good}, \text{risk}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
- \*  $\{\text{credit\_history}=\text{good}, \text{income}=>35\} \implies \{\text{risk}=\text{low}\}$ 
  - $\frac{s(\{\text{credit\_history}=\text{good}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{credit\_history}=\text{good}, \text{income}=>35\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
- 2 consequents
  - \*  $\{\text{risk}=\text{low}\} \implies \{\text{credit\_history}=\text{good}, \text{income}=>35\}$ 
    - $\frac{s(\{\text{credit\_history}=\text{good}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{risk}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
  - \*  $\{\text{income}=>35\} \implies \{\text{credit\_history}=\text{good}, \text{risk}=\text{low}\}$ 
    - $\frac{s(\{\text{credit\_history}=\text{good}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{income}=>35\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
  - \*  $\{\text{credit\_history}=\text{good}\} \implies \{\text{income}=>35, \text{risk}=\text{low}\}$ 
    - $\frac{s(\{\text{credit\_history}=\text{good}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{credit\_history}=\text{good}\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
- $\{\text{debt}=\text{low}, \text{income}=>35, \text{risk}=\text{low}\}$ 
  - 1 consequent
    - \*  $\{\text{income}=>35, \text{risk}=\text{low}\} \implies \{\text{debt}=\text{low}\}$ 
      - $\frac{s(\{\text{debt}=\text{low}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{income}=>35, \text{risk}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
    - \*  $\{\text{debt}=\text{low}, \text{risk}=\text{low}\} \implies \{\text{income}=>35\}$ 
      - $\frac{s(\{\text{debt}=\text{low}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{debt}=\text{low}, \text{risk}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
    - \*  $\{\text{debt}=\text{low}, \text{income}=>35\} \implies \{\text{risk}=\text{low}\}$ 
      - $\frac{s(\{\text{debt}=\text{low}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{debt}=\text{low}, \text{income}=>35\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9$
  - 2 consequents
    - \*  $\{\text{risk}=\text{low}\} \implies \{\text{debt}=\text{low}, \text{income}=>35\}$ 
      - $\frac{s(\{\text{debt}=\text{low}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{risk}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
    - \*  $\{\text{income}=>35\} \implies \{\text{debt}=\text{low}, \text{risk}=\text{low}\}$ 
      - $\frac{s(\{\text{debt}=\text{low}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{income}=>35\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
    - \*  $\{\text{debt}=\text{low}\} \implies \{\text{income}=>35, \text{risk}=\text{low}\}$ 
      - $\frac{s(\{\text{debt}=\text{low}, \text{income}=>35, \text{risk}=\text{low}\})}{s(\{\text{debt}=\text{low}\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
- $\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{income}=15-35\}$ 
  - 1 consequent
    - \*  $\{\text{collateral}=\text{none}, \text{income}=15-35\} \implies \{\text{debt}=\text{high}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{income}=15-35\})}{s(\{\text{collateral}=\text{none}, \text{income}=15-35\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
    - \*  $\{\text{debt}=\text{high}, \text{income}=15-35\} \implies \{\text{collateral}=\text{none}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{income}=15-35\})}{s(\{\text{debt}=\text{high}, \text{income}=15-35\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$

- \*  $\{\text{debt}=\text{high}, \text{collateral}=\text{none}\} \implies \{\text{income}=\text{15-35}\}$ 
  - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{income}=\text{15-35}\})}{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}\})} = \frac{\frac{3}{14}}{\frac{6}{14}} = \frac{3}{6} < 0.9$
- 2 consequents
  - \*  $\{\text{income}=\text{15-35}\} \implies \{\text{debt}=\text{high}, \text{collateral}=\text{none}\}$ 
    - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{income}=\text{15-35}\})}{s(\{\text{income}=\text{15-35}\})} = \frac{\frac{3}{14}}{\frac{3}{5}} = \frac{3}{5} < 0.9$
  - \*  $\{\text{collateral}=\text{none}\} \implies \{\text{debt}=\text{high}, \text{income}=\text{15-35}\}$ 
    - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{income}=\text{15-35}\})}{s(\{\text{collateral}=\text{none}\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9$
  - \*  $\{\text{debt}=\text{high}\} \implies \{\text{collateral}=\text{none}, \text{income}=\text{15-35}\}$ 
    - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{income}=\text{15-35}\})}{s(\{\text{debt}=\text{high}\})} = \frac{\frac{3}{14}}{\frac{7}{14}} = \frac{3}{7} < 0.9$
- $\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{risk}=\text{high}\}$ 
  - 1 consequent
    - \*  $\{\text{collateral}=\text{none}, \text{risk}=\text{high}\} \implies \{\text{debt}=\text{high}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{risk}=\text{high}\})}{s(\{\text{collateral}=\text{none}, \text{risk}=\text{high}\})} = \frac{\frac{4}{14}}{\frac{6}{14}} = \frac{4}{6} < 0.9$
    - \*  $\{\text{debt}=\text{high}, \text{risk}=\text{high}\} \implies \{\text{collateral}=\text{none}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{risk}=\text{high}\})}{s(\{\text{debt}=\text{high}, \text{risk}=\text{high}\})} = \frac{\frac{4}{14}}{\frac{14}{14}} = 1 \geq 0.9$
    - \*  $\{\text{debt}=\text{high}, \text{collateral}=\text{none}\} \implies \{\text{risk}=\text{high}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{risk}=\text{high}\})}{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}\})} = \frac{\frac{4}{14}}{\frac{6}{14}} = \frac{4}{6} < 0.9$
  - 2 consequents
    - \*  $\{\text{risk}=\text{high}\} \implies \{\text{debt}=\text{high}, \text{collateral}=\text{none}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{risk}=\text{high}\})}{s(\{\text{risk}=\text{high}\})} = \frac{\frac{4}{14}}{\frac{5}{14}} = \frac{4}{5} < 0.9$
    - \*  $\{\text{collateral}=\text{none}\} \implies \{\text{debt}=\text{high}, \text{risk}=\text{high}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{risk}=\text{high}\})}{s(\{\text{collateral}=\text{none}\})} = \frac{\frac{4}{14}}{\frac{11}{14}} = \frac{4}{11} < 0.9$
    - \*  $\{\text{debt}=\text{high}\} \implies \{\text{collateral}=\text{none}, \text{risk}=\text{high}\}$ 
      - $\frac{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}, \text{risk}=\text{high}\})}{s(\{\text{debt}=\text{high}\})} = \frac{\frac{4}{14}}{\frac{7}{14}} = \frac{4}{7} < 0.9$
- $\{\text{collateral}=\text{none}, \text{income}=\text{0-15}, \text{risk}=\text{high}\}$ 
  - 1 consequent
    - \*  $\{\text{income}=\text{0-15}, \text{risk}=\text{high}\} \implies \{\text{collateral}=\text{none}\}$ 
      - $\frac{s(\{\text{collateral}=\text{none}, \text{income}=\text{0-15}, \text{risk}=\text{high}\})}{s(\{\text{income}=\text{0-15}, \text{risk}=\text{high}\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
    - \*  $\{\text{collateral}=\text{none}, \text{risk}=\text{high}\} \implies \{\text{income}=\text{0-15}\}$ 
      - $\frac{s(\{\text{collateral}=\text{none}, \text{income}=\text{0-15}, \text{risk}=\text{high}\})}{s(\{\text{collateral}=\text{none}, \text{risk}=\text{high}\})} = \frac{\frac{3}{14}}{\frac{5}{14}} = \frac{3}{5} < 0.9$
    - \*  $\{\text{collateral}=\text{none}, \text{income}=\text{0-15}\} \implies \{\text{risk}=\text{high}\}$ 
      - $\frac{s(\{\text{collateral}=\text{none}, \text{income}=\text{0-15}, \text{risk}=\text{high}\})}{s(\{\text{collateral}=\text{none}, \text{income}=\text{0-15}\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9$
  - 2 consequents
    - \*  $\{\text{risk}=\text{high}\} \implies \{\text{collateral}=\text{none}, \text{income}=\text{0-15}\}$

$$\begin{aligned}
 & \cdot \frac{s(\{collateral=none, income=0-15, risk=high\})}{s(\{risk=high\})} = \frac{\frac{3}{14}}{\frac{8}{14}} = \frac{3}{8} < 0.9 \\
 * \{income=0-15\} & \implies \{collateral=none, risk=high\} \\
 & \cdot \frac{s(\{collateral=none, income=0-15, risk=high\})}{s(\{income=0-15\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9 \\
 * \{collateral=none\} & \implies \{income=0-15, risk=high\} \\
 & \cdot \frac{s(\{collateral=none, income=0-15, risk=high\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9
 \end{aligned}$$

- {collateral=none, income=15-35, risk=moderate}

– 1 consequent

$$\begin{aligned}
 * \{income=15-35, risk=moderate\} & \implies \{collateral=none\} \\
 & \cdot \frac{s(\{collateral=none, income=15-35, risk=moderate\})}{s(\{income=15-35, risk=moderate\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9 \\
 * \{collateral=none, risk=moderate\} & \implies \{income=15-35\} \\
 & \cdot \frac{s(\{collateral=none, income=15-35, risk=moderate\})}{s(\{collateral=none, risk=moderate\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9 \\
 * \{collateral=none, income=15-35\} & \implies \{risk=moderate\} \\
 & \cdot \frac{s(\{collateral=none, income=15-35, risk=moderate\})}{s(\{collateral=none, income=15-35\})} = \frac{\frac{3}{14}}{\frac{8}{14}} = \frac{3}{8} < 0.9
 \end{aligned}$$

– 2 consequents

$$\begin{aligned}
 * \{risk=moderate\} & \implies \{collateral=none, income=15-35\} \\
 & \cdot \frac{s(\{collateral=none, income=15-35, risk=moderate\})}{s(\{risk=moderate\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9 \\
 * \{income=15-35\} & \implies \{collateral=none, risk=moderate\} \\
 & \cdot \frac{s(\{collateral=none, income=15-35, risk=moderate\})}{s(\{income=15-35\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = \frac{3}{3} < 0.9 \\
 * \{collateral=none\} & \implies \{income=15-35, risk=moderate\} \\
 & \cdot \frac{s(\{collateral=none, income=15-35, risk=moderate\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9
 \end{aligned}$$

- {collateral=none, income=>35, risk=low}

– 1 consequent

$$\begin{aligned}
 * \{income=>35, risk=low\} & \implies \{collateral=none\} \\
 & \cdot \frac{s(\{collateral=none, income=>35, risk=low\})}{s(\{income=>35, risk=low\})} = \frac{\frac{3}{14}}{\frac{8}{14}} = \frac{3}{8} < 0.9 \\
 * \{collateral=none, risk=low\} & \implies \{income=>35\} \\
 & \cdot \frac{s(\{collateral=none, income=>35, risk=low\})}{s(\{collateral=none, risk=low\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9 \\
 * \{collateral=none, income=>35\} & \implies \{risk=low\} \\
 & \cdot \frac{s(\{collateral=none, income=>35, risk=low\})}{s(\{collateral=none, income=>35\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = 1 \geq 0.9
 \end{aligned}$$

– 2 consequents

$$\begin{aligned}
 * \{risk=low\} & \implies \{collateral=none, income=>35\} \\
 & \cdot \frac{s(\{collateral=none, income=>35, risk=low\})}{s(\{risk=low\})} = \frac{\frac{3}{14}}{\frac{4}{14}} = \frac{3}{4} < 0.9 \\
 * \{income=>35\} & \implies \{collateral=none, risk=low\} \\
 & \cdot \frac{s(\{collateral=none, income=>35, risk=low\})}{s(\{income=>35\})} = \frac{\frac{3}{14}}{\frac{8}{14}} = \frac{3}{8} < 0.9 \\
 * \{collateral=none\} & \implies \{income=>35, risk=low\} \\
 & \cdot \frac{s(\{collateral=none, income=>35, risk=low\})}{s(\{collateral=none\})} = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} < 0.9
 \end{aligned}$$

### 3.2.3.4 Totals

Together with the previous rules, the ones just generated make a total of list of the following 18 rules. Note that each one of them achieved a confidence of 100% (which is of course  $\geq 90\%$ ). Note that these were varified with WEKA. WEKA seemed to have produced these except in a different order.

1.  $\{\text{income}=0-15\} \implies \{\text{collateral}=\text{none}\}$
2.  $\{\text{income}=15-35\} \implies \{\text{collateral}=\text{none}\}$
3.  $\{\text{risk}=\text{high}\} \implies \{\text{collateral}=\text{none}\}$
4.  $\{\text{collateral}=\text{adequate}\} \implies \{\text{income}=>35\}$
5.  $\{\text{income}=0-15\} \implies \{\text{risk}=\text{high}\}$
6.  $\{\text{risk}=\text{low}\} \implies \{\text{income}=>35\}$
7.  $\{\text{credit\_history}=\text{good}, \text{risk}=\text{low}\} \implies \{\text{income}=>35\}$
8.  $\{\text{credit\_history}=\text{good}, \text{income}=>35\} \implies \{\text{risk}=\text{low}\}$
9.  $\{\text{deb}=\text{low}, \text{risk}=\text{low}\} \implies \{\text{income}=>35\}$
10.  $\{\text{debt}=\text{high}, \text{income}=15-35\} \implies \{\text{collateral}=\text{none}\}$
11.  $\{\text{debt}=\text{high}, \text{risk}=\text{high}\} \implies \{\text{collateral}=\text{none}\}$
12.  $\{\text{income}=0-15, \text{risk}=\text{high}\} \implies \{\text{collateral}=\text{none}\}$
13.  $\{\text{collateral}=\text{none}, \text{income}=0-15\} \implies \{\text{risk}=\text{high}\}$
14.  $\{\text{income}=0-15\} \implies \{\text{collateral}=\text{none}, \text{risk}=\text{high}\}$
15.  $\{\text{income}=15-35, \text{risk}=\text{moderate}\} \implies \{\text{collateral}=\text{none}\}$
16.  $\{\text{collateral}=\text{none}, \text{risk}=\text{moderate}\} \implies \{\text{income}=15-35\}$
17.  $\{\text{collateral}=\text{none}, \text{risk}=\text{low}\} \implies \{\text{income}=>35\}$
18.  $\{\text{collateral}=\text{none}, \text{income}=>35\} \implies \{\text{risk}=\text{low}\}$