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

Piotr Mardziel

14th April 2004
Contents

3 Project 3: Data Pre-processing, Mining, and Evaluation of Association Rules: Homework ................................. 2
  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

<table>
<thead>
<tr>
<th>GROUP MEMBER</th>
<th>USERNAME</th>
<th>TASKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piotr Mardziel</td>
<td>piotrm</td>
<td>everything</td>
</tr>
</tbody>
</table>

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=low}
- {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}
\[ \begin{align*}
\{ & \text{credit\_history=good, debt=high, collateral=none, income=0-15, risk=high} \} \\
\{ & \text{credit\_history=good, debt=high, collateral=none, income=15-35, risk=moderate} \} \\
\{ & \text{credit\_history=good, debt=high, collateral=none, income=>35, risk=low} \} \\
\{ & \text{credit\_history=bad, debt=high, collateral=none, income=15-35, risk=high} \} \\
\end{align*} \]

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.

\[ \begin{align*}
\{ & \text{credit\_history=bad} \} \\
\{ & \text{credit\_history=unknown} \} \\
\{ & \text{credit\_history=good} \} \\
\{ & \text{debt=low} \} \\
\{ & \text{debt=high} \} \\
\{ & \text{collateral=none} \} \\
\{ & \text{collateral=adequate} \} \\
\{ & \text{income=0-15} \} \\
\{ & \text{income=15-35} \} \\
\{ & \text{income=>35} \} \\
\{ & \text{risk=low} \} \\
\{ & \text{risk=moderate} \} \\
\{ & \text{risk=high} \} \\
\end{align*} \]

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.

<table>
<thead>
<tr>
<th>CANDIDATE</th>
<th>SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>{credit_history=bad}</td>
<td>4/14</td>
</tr>
<tr>
<td>{credit_history=unknown}</td>
<td>5/14</td>
</tr>
<tr>
<td>{credit_history=good}</td>
<td>5/14</td>
</tr>
<tr>
<td>{debt=low}</td>
<td>7/14</td>
</tr>
<tr>
<td>{debt=high}</td>
<td>7/14</td>
</tr>
<tr>
<td>{collateral=none}</td>
<td>11/14</td>
</tr>
<tr>
<td>{collateral=adequate}</td>
<td>3/14</td>
</tr>
<tr>
<td>{income=0-15}</td>
<td>3/14</td>
</tr>
<tr>
<td>{income=15-35}</td>
<td>5/14</td>
</tr>
<tr>
<td>{income=&gt;35}</td>
<td>6/14</td>
</tr>
<tr>
<td>{risk=low}</td>
<td>5/14</td>
</tr>
<tr>
<td>{risk=moderate}</td>
<td>4/14</td>
</tr>
<tr>
<td>{risk=high}</td>
<td>5/14</td>
</tr>
</tbody>
</table>
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 if 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}`
• \{\text{collateral}=\text{none}, \text{risk}=\text{high}\}
• \{\text{collateral}=\text{adequate}, \text{income}=0\text{-}15\}
• \{\text{collateral}=\text{adequate}, \text{income}=15\text{-}35\}
• \{\text{collateral}=\text{adequate}, \text{income}=>35\}
• \{\text{collateral}=\text{adequate}, \text{risk}=\text{low}\}
• \{\text{collateral}=\text{adequate}, \text{risk}=\text{moderate}\}
• \{\text{collateral}=\text{adequate}, \text{risk}=\text{high}\}
• \{\text{income}=0\text{-}15, \text{income}=15\text{-}35\}
• \{\text{income}=0\text{-}15, \text{income}=>35\}
• \{\text{income}=0\text{-}15, \text{risk}=\text{low}\}
• \{\text{income}=0\text{-}15, \text{risk}=\text{moderate}\}
• \{\text{income}=0\text{-}15, \text{risk}=\text{high}\}
• \{\text{income}=15\text{-}35, \text{income}=>35\}
• \{\text{income}=15\text{-}35, \text{risk}=\text{low}\}
• \{\text{income}=15\text{-}35, \text{risk}=\text{moderate}\}
• \{\text{income}=15\text{-}35, \text{risk}=\text{high}\}
• \{\text{income}=>35, \text{risk}=\text{low}\}
• \{\text{income}=>35, \text{risk}=\text{moderate}\}
• \{\text{income}=>35, \text{risk}=\text{high}\}
• \{\text{risk}=\text{low}, \text{risk}=\text{moderate}\}
• \{\text{risk}=\text{low}, \text{risk}=\text{high}\}
• \{\text{risk}=\text{moderate}, \text{risk}=\text{high}\}

\subsection*{3.2.2.2 \textit{F}_2} In order to prune a lot of these, their support is computed.

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

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

<table>
<thead>
<tr>
<th>CANDIDATE</th>
<th>SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>{credit_history=bad, debt=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{credit_history=bad, collateral=none}</td>
<td>3/14</td>
</tr>
<tr>
<td>{credit_history=unknown, debt=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{credit_history=unknown, collateral=none}</td>
<td>4/14</td>
</tr>
<tr>
<td>{credit_history=good, debt=high}</td>
<td>4/14</td>
</tr>
<tr>
<td>{credit_history=good, collateral=none}</td>
<td>4/14</td>
</tr>
<tr>
<td>{credit_history=good, income=&gt;35}</td>
<td>3/14</td>
</tr>
<tr>
<td>{credit_history=good, risk=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=low, collateral=none}</td>
<td>5/14</td>
</tr>
<tr>
<td>{debt=low, income=&gt;35}</td>
<td>4/14</td>
</tr>
<tr>
<td>{debt=low, risk=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=low, risk=moderate}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=high, collateral=none}</td>
<td>6/14</td>
</tr>
<tr>
<td>{debt=high, income=15-35}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=high, risk=high}</td>
<td>4/14</td>
</tr>
<tr>
<td>{collateral=none, income=0-15}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, income=15-35}</td>
<td>5/14</td>
</tr>
<tr>
<td>{collateral=none, income=&gt;35}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, risk=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, risk=moderate}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, risk=high}</td>
<td>5/14</td>
</tr>
<tr>
<td>{collateral=adequate, income=&gt;35}</td>
<td>3/14</td>
</tr>
<tr>
<td>{income=0-15, risk=high}</td>
<td>3/14</td>
</tr>
<tr>
<td>{income=15-35, risk=moderate}</td>
<td>3/14</td>
</tr>
</tbody>
</table>
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}
- {deb=low, collateral=none, income=>35}
- {deb=low, collateral=none, risk=low}
- {deb=low, collateral=none, risk=moderate}
- {deb=low, income=>35, risk=low}
- * {deb=low, income=>35, risk=moderate}
- * {deb=low, risk=low, risk=moderate}
- {deb=high, collateral=none, income=15-35}
- {deb=high, collateral=none, risk=high}
- * {deb=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\}
• \{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.

<table>
<thead>
<tr>
<th>CANDIDATE</th>
<th>SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>{credit_history=bad, debt=low, collateral=none}</td>
<td>2/14</td>
</tr>
<tr>
<td>{credit_history=unknown, debt=low, collateral=none}</td>
<td>2/14</td>
</tr>
<tr>
<td>{credit_history=good, debt=high, collateral=none}</td>
<td>3/14</td>
</tr>
<tr>
<td>{credit_history=good, collateral=none, income=&gt;35}</td>
<td>2/14</td>
</tr>
<tr>
<td>{credit_history=good, collateral=none, risk=low}</td>
<td>2/14</td>
</tr>
<tr>
<td>{credit_history=good, income=&gt;35, risk=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=low, collateral=none, income=&gt;35}</td>
<td>2/14</td>
</tr>
<tr>
<td>{debt=low, collateral=none, risk=low}</td>
<td>2/14</td>
</tr>
<tr>
<td>{debt=low, collateral=none, risk=moderate}</td>
<td>2/14</td>
</tr>
<tr>
<td>CANDIDATE</td>
<td>SUPPORT</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>{debt=low, income=&gt;35, risk=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=high, collateral=none, income=15-35}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=high, collateral=none, risk=high}</td>
<td>4/14</td>
</tr>
<tr>
<td>{collateral=none, income=0-15, risk=high}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, income=15-35, risk=moderate}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, income=&gt;35, risk=low}</td>
<td>3/14</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>CANDIDATE</th>
<th>SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>{credit history=good, debt=high, collateral=none}</td>
<td>3/14</td>
</tr>
<tr>
<td>{credit history=good, income=&gt;35, risk=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=low, income=&gt;35, risk=low}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=high, collateral=none, income=15-35}</td>
<td>3/14</td>
</tr>
<tr>
<td>{debt=high, collateral=none, risk=high}</td>
<td>4/14</td>
</tr>
<tr>
<td>{collateral=none, income=0-15, risk=high}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, income=15-35, risk=moderate}</td>
<td>3/14</td>
</tr>
<tr>
<td>{collateral=none, income=&gt;35, risk=low}</td>
<td>3/14</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>CANDIDATE</th>
<th>SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>{credit history=bad}</td>
<td>4/14</td>
</tr>
<tr>
<td>{credit history=unknown}</td>
<td>5/14</td>
</tr>
<tr>
<td>{credit history=good}</td>
<td>5/14</td>
</tr>
<tr>
<td>{debt=low}</td>
<td>7/14</td>
</tr>
<tr>
<td>{debt=high}</td>
<td>7/14</td>
</tr>
<tr>
<td>{collateral=none}</td>
<td>11/14</td>
</tr>
<tr>
<td>{collateral=adequate}</td>
<td>3/14</td>
</tr>
<tr>
<td>{income=0-15}</td>
<td>3/14</td>
</tr>
<tr>
<td>{income=15-35}</td>
<td>5/14</td>
</tr>
<tr>
<td>{income=&gt;35}</td>
<td>6/14</td>
</tr>
<tr>
<td>{risk=low}</td>
<td>5/14</td>
</tr>
<tr>
<td>{risk=moderate}</td>
<td>4/14</td>
</tr>
<tr>
<td>{risk=high}</td>
<td>5/14</td>
</tr>
</tbody>
</table>
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

- \{\text{credit\_history}=\text{bad}, \text{debt}=\text{low}\} - \{\text{debt}=\text{low}\} \Rightarrow \{\text{credit\_history}=\text{bad}\}
  
  \[ \frac{s(\{\text{credit\_history}=\text{bad}, \text{debt}=\text{low}\})}{s(\{\text{credit\_history}=\text{bad}\})} = \frac{\frac{3}{14}}{\frac{3}{14}} = \frac{3}{3} < 0.9 \]
- \{\text{credit\_history}=\text{bad}\} \Rightarrow \{\text{debt}=\text{low}\}
\[
s\left(\{\text{credit\_history=bad, debt=low}\}\right) = \frac{4}{11} = \frac{3}{9} < 0.9
\]

- \{credit\_history=bad, collateral=none\}
  - \{collateral=none\} \implies \{credit\_history=bad\}
    \[
    s\left(\{\text{credit\_history=bad, collateral=none}\}\right) = \frac{3}{4} < 0.9
    \]
  - \{credit\_history=bad\} \implies \{collateral=none\}
    \[
    s\left(\{\text{credit\_history=bad, collateral=none}\}\right) = \frac{3}{4} < 0.9
    \]

- \{credit\_history=unknown, debt=low\}
  - \{debt=low\} \implies \{credit\_history=unknown\}
    \[
    s\left(\{\text{credit\_history=unknown, debt=low}\}\right) = \frac{3}{9} = \frac{3}{9} < 0.9
    \]
  - \{credit\_history=unknown\} \implies \{debt=low\}
    \[
    s\left(\{\text{credit\_history=unknown, debt=low}\}\right) = \frac{3}{9} = \frac{3}{9} < 0.9
    \]

- \{credit\_history=unknown, collateral=none\}
  - \{collateral=none\} \implies \{credit\_history=unknown\}
    \[
    s\left(\{\text{credit\_history=unknown, collateral=none}\}\right) = \frac{4}{11} = \frac{4}{11} < 0.9
    \]
  - \{credit\_history=unknown\} \implies \{collateral=none\}
    \[
    s\left(\{\text{credit\_history=unknown, collateral=none}\}\right) = \frac{4}{11} = \frac{4}{11} < 0.9
    \]

- \{credit\_history=good, debt=high\}
  - \{debt=high\} \implies \{credit\_history=good\}
    \[
    s\left(\{\text{credit\_history=good, debt=high}\}\right) = \frac{4}{9} = \frac{4}{9} < 0.9
    \]
  - \{credit\_history=good\} \implies \{debt=high\}
    \[
    s\left(\{\text{credit\_history=good, debt=high}\}\right) = \frac{4}{9} = \frac{4}{9} < 0.9
    \]

- \{credit\_history=good, collateral=none\}
  - \{collateral=none\} \implies \{credit\_history=good\}
    \[
    s\left(\{\text{credit\_history=good, collateral=none}\}\right) = \frac{4}{11} = \frac{4}{11} < 0.9
    \]
  - \{credit\_history=good\} \implies \{collateral=none\}
    \[
    s\left(\{\text{credit\_history=good, collateral=none}\}\right) = \frac{4}{11} = \frac{4}{11} < 0.9
    \]

- \{credit\_history=good, income=>35\}
  - \{income=>35\} \implies \{credit\_history=good\}
    \[
    s\left(\{\text{credit\_history=good, income=>35}\}\right) = \frac{3}{6} = \frac{3}{6} < 0.9
    \]
\[- \{\text{credit\_history}=\text{good}\} \implies \{\text{income}>&35\}\]
\[s(\{\text{credit\_history}=\text{good},\text{income}>&35\}) = \frac{1}{2} < 0.9\]
\[s(\{\text{credit\_history}=\text{good}\}) = \frac{1}{2} < 0.9\]

- \{\text{credit\_history}=\text{good}, \text{risk}=\text{low}\}

- \{\text{risk}=\text{low}\} \implies \{\text{credit\_history}=\text{good}\}

\[s(\{\text{credit\_history}=\text{good},\text{risk}=\text{low}\}) = \frac{3}{5} < 0.9\]

- \{\text{credit\_history}=\text{good}\} \implies \{\text{risk}=\text{low}\}

\[s(\{\text{credit\_history}=\text{good}\}) = \frac{3}{5} < 0.9\]

- \{\text{debt}=\text{low}, \text{collateral}=\text{none}\}

- \{\text{collateral}=\text{none}\} \implies \{\text{debt}=\text{low}\}

\[s(\{\text{debt}=\text{low},\text{collateral}=\text{none}\}) = \frac{5}{9} = 0.5 < 0.9\]

- \{\text{debt}=\text{low}\} \implies \{\text{collateral}=\text{none}\}

\[s(\{\text{debt}=\text{low}\}) = \frac{5}{9} < 0.9\]

- \{\text{debt}=\text{low}, \text{income}>&35\}

- \{\text{income}>&35\} \implies \{\text{debt}=\text{low}\}

\[s(\{\text{debt}=\text{low},\text{income}>&35\}) = \frac{4}{6} = 0.6 < 0.9\]

- \{\text{debt}=\text{low}\} \implies \{\text{income}>&35\}

\[s(\{\text{debt}=\text{low}\}) = \frac{4}{6} < 0.9\]

- \{\text{debt}=\text{low}, \text{risk}=\text{low}\}

- \{\text{risk}=\text{low}\} \implies \{\text{debt}=\text{low}\}

\[s(\{\text{debt}=\text{low},\text{risk}=\text{low}\}) = \frac{3}{5} < 0.9\]

- \{\text{debt}=\text{low}\} \implies \{\text{risk}=\text{low}\}

\[s(\{\text{debt}=\text{low}\}) = \frac{3}{5} < 0.9\]

- \{\text{debt}=\text{low}, \text{risk}=\text{moderate}\}

- \{\text{risk}=\text{moderate}\} \implies \{\text{debt}=\text{low}\}

\[s(\{\text{debt}=\text{low},\text{risk}=\text{moderate}\}) = \frac{3}{4} < 0.9\]

- \{\text{debt}=\text{low}\} \implies \{\text{risk}=\text{moderate}\}

\[s(\{\text{debt}=\text{low}\}) = \frac{3}{4} < 0.9\]

- \{\text{debt}=\text{high}, \text{collateral}=\text{none}\}

- \{\text{collateral}=\text{none}\} \implies \{\text{debt}=\text{high}\}
\[ s(\text{debt=high,collateral=none}) \]
\[ \frac{s(\text{collateral=none})}{s(\text{collateral=none})} = \frac{\hat{p}}{\hat{p}} = \frac{6}{11} < 0.9 \]
\[ \quad \text{– \{debt=high\} \implies \{collateral=none\}} \]
\[ s(\text{debt=high,collateral=none}) \]
\[ \frac{s(\text{debt=high,collateral=none})}{s(\text{debt=high})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{debt=high, income=15-35\}} \]
\[ s(\text{income=15-35}) \]
\[ \frac{s(\text{income=15-35})}{s(\text{income=15-35})} = \frac{\hat{p}}{\hat{p}} = \frac{3}{11} < 0.9 \]
\[ \quad \text{– \{income=15-35\} \implies \{debt=high\} \]
\[ s(\text{debt=high,collateral=none}) \]
\[ \frac{s(\text{debt=high,collateral=none})}{s(\text{debt=high})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{debt=high, risk=high\}} \]
\[ s(\text{risk=high}) \]
\[ \frac{s(\text{risk=high})}{s(\text{risk=high})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{risk=high\} \implies \{debt=high\} \]
\[ s(\text{debt=high,collateral=none}) \]
\[ \frac{s(\text{debt=high,collateral=none})}{s(\text{debt=high})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{collateral=none, income=0-15\}} \]
\[ s(\text{income=0-15}) \]
\[ \frac{s(\text{income=0-15})}{s(\text{income=0-15})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{income=0-15\} \implies \{collateral=none\} \]
\[ s(\text{collateral=none,collateral=none}) \]
\[ \frac{s(\text{collateral=none,collateral=none})}{s(\text{collateral=none})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{collateral=none, income=15-35\}} \]
\[ s(\text{income=15-35}) \]
\[ \frac{s(\text{income=15-35})}{s(\text{income=15-35})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{income=15-35\} \implies \{collateral=none\} \]
\[ s(\text{collateral=none,collateral=none}) \]
\[ \frac{s(\text{collateral=none,collateral=none})}{s(\text{collateral=none})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{collateral=none, income=>35\}} \]
\[ s(\text{income=>35}) \]
\[ \frac{s(\text{income=>35})}{s(\text{income=>35})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{income=>35\} \implies \{collateral=none\} \]
\[ s(\text{collateral=none,collateral=none}) \]
\[ \frac{s(\text{collateral=none,collateral=none})}{s(\text{collateral=none})} = \frac{\hat{p}}{\hat{p}} = \frac{\hat{p}}{\hat{p}} < 0.9 \]
\[ \quad \text{– \{collateral=none, risk=low\}} \]
- \{\text{risk}=\text{low}\} \implies \{\text{collateral}=\text{none}\}
  \* \frac{s(\text{collateral}=\text{none}, \text{risk}=\text{low})}{s(\text{risk}=\text{low})} = \frac{2}{11} = \frac{3}{5} < 0.9

- \{\text{collateral}=\text{none}\} \implies \{\text{risk}=\text{low}\}
  \* \frac{s(\text{collateral}=\text{none}, \text{risk}=\text{low})}{s(\text{collateral}=\text{none})} = \frac{2}{11} = \frac{3}{5} < 0.9

- \{\text{collateral}=\text{none}, \text{risk}=\text{moderate}\}
  \* \frac{s(\text{collateral}=\text{none}, \text{risk}=\text{moderate})}{s(\text{risk}=\text{moderate})} = \frac{3}{11} = \frac{3}{4} < 0.9

- \{\text{collateral}=\text{none}, \text{risk}=\text{high}\}
  \* \frac{s(\text{collateral}=\text{none}, \text{risk}=\text{high})}{s(\text{risk}=\text{high})} = \frac{5}{11} = \frac{5}{6} < 0.9

- \{\text{collateral}=\text{adequate}, \text{income}=\text{>35}\}
  \* \frac{s(\text{collateral}=\text{adequate}, \text{income}=\text{>35})}{s(\text{income}=\text{>35})} = \frac{4}{9} = \frac{3}{5} < 0.9

- \{\text{collateral}=\text{adequate}\} \implies \{\text{income}=\text{>35}\}
  \* \frac{s(\text{collateral}=\text{adequate}, \text{income}=\text{>35})}{s(\text{collateral}=\text{adequate})} = \frac{4}{9} = \frac{5}{6} < 0.9

- \{\text{income}=0-15, \text{risk}=\text{high}\}
  \* \frac{s(\text{income}=0-15, \text{risk}=\text{high})}{s(\text{risk}=\text{high})} = \frac{3}{4} = \frac{3}{5} < 0.9

- \{\text{income}=0-15\} \implies \{\text{risk}=\text{high}\}
  \* \frac{s(\text{income}=0-15, \text{risk}=\text{high})}{s(\text{income}=0-15)} = \frac{3}{4} = \frac{1}{2} < 0.9

- \{\text{income}=15-35, \text{risk}=\text{moderate}\}
  \* \frac{s(\text{income}=15-35, \text{risk}=\text{moderate})}{s(\text{risk}=\text{moderate})} = \frac{3}{4} = \frac{3}{5} < 0.9

- \{\text{income}=15-35\} \implies \{\text{risk}=\text{moderate}\}
  \* \frac{s(\text{income}=15-35, \text{risk}=\text{moderate})}{s(\text{income}=15-35)} = \frac{3}{4} = \frac{3}{5} < 0.9

- \{\text{income}=\text{>35}, \text{risk}=\text{low}\}
\[ \{\text{risk}=\text{low}\} \implies \{\text{income}>35\} \]
\[ \frac{s(\{\text{income}>35, \text{risk}=\text{low}\})}{s(\{\text{risk}=\text{low}\})} = \frac{5}{6} < 0.9 \]

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 \{\text{collateral}=\text{none}\}
2. \{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\}

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.

- \{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}, \text{collateral}=\text{none}\}
  - 1 consequent
    * \{\text{debt}=\text{high}, \text{collateral}=\text{none}\} \implies \{\text{credit}_\text{history}=\text{good}\}
      \[ \frac{s(\{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}, \text{collateral}=\text{none}\})}{s(\{\text{debt}=\text{high}, \text{collateral}=\text{none}\})} = \frac{5}{11} = \frac{5}{6} < 0.9 \]
    * \{\text{credit}_\text{history}=\text{good}, \text{collateral}=\text{none}\} \implies \{\text{debt}=\text{high}\}
      \[ \frac{s(\{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}, \text{collateral}=\text{none}\})}{s(\{\text{debt}=\text{high}\})} = \frac{1}{4} = \frac{1}{4} < 0.9 \]
    * \{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}\} \implies \{\text{collateral}=\text{none}\}
      \[ \frac{s(\{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}, \text{collateral}=\text{none}\})}{s(\{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}\})} = \frac{3}{4} = \frac{3}{4} < 0.9 \]
  - 2 consequents
    * \{\text{collateral}=\text{none}\} \implies \{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}\}
      \[ \frac{s(\{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}, \text{collateral}=\text{none}\})}{s(\{\text{collateral}=\text{none}\})} = \frac{3}{4} = \frac{3}{4} < 0.9 \]
    * \{\text{debt}=\text{high}\} \implies \{\text{credit}_\text{history}=\text{good}, \text{collateral}=\text{none}\}
      \[ \frac{s(\{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}, \text{collateral}=\text{none}\})}{s(\{\text{debt}=\text{high}\})} = \frac{1}{4} = \frac{1}{4} < 0.9 \]
    * \{\text{credit}_\text{history}=\text{good}\} \implies \{\text{debt}=\text{high}, \text{collateral}=\text{none}\}
      \[ \frac{s(\{\text{credit}_\text{history}=\text{good}, \text{debt}=\text{high}, \text{collateral}=\text{none}\})}{s(\{\text{credit}_\text{history}=\text{good}\})} = \frac{3}{7} = \frac{3}{7} < 0.9 \]

- \{\text{credit}_\text{history}=\text{good}, \text{income}>35, \text{risk}=\text{low}\}
  - 1 consequent
\[
\begin{align*}
* \{ \text{income} \rightarrow 35, \text{risk} = \text{low} \} \implies \{ \text{credit\_history} = \text{good} \} \\
& s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{income} \geq 35, \text{risk} = \text{low} \}) = \frac{3}{15} > \frac{3}{9} < 0.9
\end{align*}
\]

\[
\begin{align*}
* \{ \text{credit\_history} = \text{good}, \text{risk} = \text{low} \} \implies \{ \text{income} \rightarrow 35 \} \\
& s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{credit\_history} = \text{good}, \text{income} \geq 35 \} \implies \{ \text{risk} = \text{low} \} \\
& s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

- 2 consequents

\[
\begin{align*}
* \{ \text{risk} = \text{low} \} \implies \{ \text{credit\_history} = \text{good}, \text{income} \geq 35 \} \\
& s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{risk} = \text{low} \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{income} \geq 35 \} \implies \{ \text{credit\_history} = \text{good}, \text{risk} = \text{low} \} \\
& s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{income} \geq 35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{credit\_history} = \text{good} \} \implies \{ \text{income} \geq 35, \text{risk} = \text{low} \} \\
& s(\{ \text{credit\_history} = \text{good}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{credit\_history} = \text{good}\}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

- 1 consequent

\[
\begin{align*}
* \{ \text{income} \geq 35, \text{risk} = \text{low} \} \implies \{ \text{debt} = \text{low} \} \\
& s(\{ \text{debt} = \text{low}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{income} \geq 35, \text{risk} = \text{low} \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{debt} = \text{low}, \text{risk} = \text{low} \} \implies \{ \text{income} \geq 35 \} \\
& s(\{ \text{debt} = \text{low}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{debt} = \text{low}, \text{risk} = \text{low} \}) = \frac{1}{15} = 0.2 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{debt} = \text{low}, \text{income} \geq 35 \} \implies \{ \text{risk} = \text{low} \} \\
& s(\{ \text{debt} = \text{low}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{debt} = \text{low}, \text{income} \geq 35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

- 2 consequents

\[
\begin{align*}
* \{ \text{risk} = \text{low} \} \implies \{ \text{debt} = \text{low}, \text{income} \geq 35 \} \\
& s(\{ \text{debt} = \text{low}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{risk} = \text{low} \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{income} \geq 35 \} \implies \{ \text{debt} = \text{low}, \text{risk} = \text{low} \} \\
& s(\{ \text{debt} = \text{low}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{income} \geq 35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{debt} = \text{low} \} \implies \{ \text{income} \geq 35, \text{risk} = \text{low} \} \\
& s(\{ \text{debt} = \text{low}, \text{income} \geq 35, \text{risk} = \text{low} \}) \quad s(\{ \text{debt} = \text{low} \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{debt} = \text{high}, \text{collateral} = \text{none}, \text{income} = 15-35 \} \\
& s(\{ \text{debt} = \text{high}, \text{collateral} = \text{none}, \text{income} = 15-35 \}) \quad s(\{ \text{debt} = \text{high}, \text{income} = 15-35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

- 1 consequent

\[
\begin{align*}
* \{ \text{collateral} = \text{none}, \text{income} = 15-35 \} \implies \{ \text{debt} = \text{high} \} \\
& s(\{ \text{debt} = \text{high}, \text{collateral} = \text{none}, \text{income} = 15-35 \}) \quad s(\{ \text{collateral} = \text{none}, \text{income} = 15-35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]

\[
\begin{align*}
* \{ \text{debt} = \text{high}, \text{income} = 15-35 \} \implies \{ \text{collateral} = \text{none} \} \\
& s(\{ \text{debt} = \text{high}, \text{collateral} = \text{none}, \text{income} = 15-35 \}) \quad s(\{ \text{debt} = \text{high}, \text{income} = 15-35 \}) = \frac{3}{15} = 0.3 > 0.2
\end{align*}
\]
* \{debt=high, collateral=none\} \implies \{income=15-35\}
  \[\frac{s(\text{debt=high, collateral=none, income=15-35})}{s(\text{debt=high, collateral=none})} = \frac{3}{11} = \frac{3}{9} < 0.9\]

- 2 consequents

* \{income=15-35\} \implies \{debt=high, collateral=none\}
  \[\frac{s(\text{debt=high, collateral=none, income=15-35})}{s(\text{income=15-35})} = \frac{3}{11} = \frac{3}{9} < 0.9\]

* \{collateral=none\} \implies \{debt=high, income=15-35\}
  \[\frac{s(\text{debt=high, collateral=none, income=15-35})}{s(\text{collateral=none})} = \frac{3}{11} = \frac{3}{11} < 0.9\]

* \{debt=high\} \implies \{collateral=none, income=15-35\}
  \[\frac{s(\text{debt=high, collateral=none, income=15-35})}{s(\text{debt=high})} = \frac{3}{11} = \frac{3}{9} < 0.9\]

- \{debt=high, collateral=none, risk=high\}

  - 1 consequent

  * \{collateral=none, risk=high\} \implies \{debt=high\}
    \[\frac{s(\text{debt=high, collateral=none, risk=high})}{s(\text{collateral=none, risk=high})} = \frac{4}{9} < 0.9\]

  * \{debt=high, risk=high\} \implies \{collateral=none\}
    \[\frac{s(\text{debt=high, collateral=none, risk=high})}{s(\text{debt=high, risk=high})} = \frac{4}{9} < 0.9\]

  * \{debt=high, collateral=none\} \implies \{risk=high\}
    \[\frac{s(\text{debt=high, collateral=none, risk=high})}{s(\text{debt=high, collateral=none})} = \frac{4}{9} < 0.9\]

  - 2 consequents

  * \{risk=high\} \implies \{debt=high, collateral=none\}
    \[\frac{s(\text{debt=high, collateral=none, risk=high})}{s(\text{risk=high})} = \frac{4}{9} < 0.9\]

  * \{collateral=none\} \implies \{debt=high, risk=high\}
    \[\frac{s(\text{debt=high, collateral=none, risk=high})}{s(\text{collateral=none})} = \frac{4}{9} < 0.9\]

  * \{debt=high\} \implies \{collateral=none, risk=high\}
    \[\frac{s(\text{debt=high, collateral=none, risk=high})}{s(\text{debt=high})} = \frac{4}{9} < 0.9\]

- \{collateral=none, income=0-15, risk=high\}

  - 1 consequent

  * \{income=0-15, risk=high\} \implies \{collateral=none\}
    \[\frac{s(\text{collateral=none, income=0-15, risk=high})}{s(\text{income=0-15, risk=high})} = \frac{5}{9} < 0.9\]

  * \{collateral=none, risk=high\} \implies \{income=0-15\}
    \[\frac{s(\text{collateral=none, income=0-15, risk=high})}{s(\text{collateral=none, risk=high})} = \frac{5}{9} < 0.9\]

  * \{collateral=none, income=0-15\} \implies \{risk=high\}
    \[\frac{s(\text{collateral=none, income=0-15, risk=high})}{s(\text{collateral=none, income=0-15})} = \frac{5}{9} < 0.9\]

  - 2 consequents

  * \{risk=high\} \implies \{collateral=none, income=0-15\}
\[
\begin{align*}
\text{s(collateral=none, income=0-15, risk=high)} & = \frac{7}{14} = \frac{3}{6} < 0.9 \\
\text{s(collateral=none, income=0-15, risk=high)} & = \frac{1}{7} = 1 \geq 0.9 \\
\text{s(collateral=none, income=0-15, risk=high)} & = \frac{3}{11} = \frac{3}{11} < 0.9 \\
\cdot \{\text{collateral=none, income=15-35, risk=moderate}\} \\
\quad - 1 \text{ consequence} \\
\quad \cdot \{\text{income=15-35, risk=moderate} \Rightarrow \{\text{collateral=none}\} \\
\quad \quad \cdot \text{s(collateral=none, income=15-35, risk=moderate)} = \frac{7}{14} = \frac{3}{4} < 0.9 \\
\quad \cdot \{\text{collateral=none, risk=moderate} \Rightarrow \{\text{income=15-35}\} \\
\quad \quad \cdot \text{s(collateral=none, income=15-35, risk=moderate)} = \frac{1}{7} = 1 \geq 0.9 \\
\quad \cdot \{\text{collateral=none, income=15-35} \Rightarrow \{\text{risk=moderate}\} \\
\quad \quad \cdot \text{s(collateral=none, income=15-35, risk=moderate)} = \frac{3}{11} = \frac{3}{11} < 0.9 \\
\quad - 2 \text{ consequences} \\
\quad \cdot \{\text{risk=moderate} \Rightarrow \{\text{collateral=none, income=15-35}\} \\
\quad \quad \cdot \text{s(collateral=none, income=15-35, risk=moderate)} = \frac{7}{14} = \frac{3}{4} < 0.9 \\
\quad \cdot \{\text{income=15-35} \Rightarrow \{\text{collateral=none, risk=moderate}\} \\
\quad \quad \cdot \text{s(collateral=none, income=15-35, risk=moderate)} = \frac{1}{7} = 1 \geq 0.9 \\
\quad \cdot \{\text{collateral=none} \Rightarrow \{\text{income=15-35, risk=moderate}\} \\
\quad \quad \cdot \text{s(collateral=none, income=15-35, risk=moderate)} = \frac{3}{11} = \frac{3}{11} < 0.9 \\
\quad \cdot \{\text{collateral=none, income=>35, risk=low}\} \\
\quad \quad - 1 \text{ consequence} \\
\quad \quad \cdot \{\text{income=>35, risk=low} \Rightarrow \{\text{collateral=none}\} \\
\quad \quad \quad \text{s(collateral=none, income=>35, risk=low)} = \frac{7}{14} = \frac{3}{9} < 0.9 \\
\quad \quad \cdot \{\text{collateral=none, risk=low} \Rightarrow \{\text{income=>35}\} \\
\quad \quad \quad \text{s(collateral=none, income=>35, risk=low)} = \frac{1}{7} = 1 \geq 0.9 \\
\quad \quad \cdot \{\text{collateral=none} \Rightarrow \{\text{income=>35, risk=low}\} \\
\quad \quad \quad \text{s(collateral=none, income=>35, risk=low)} = \frac{3}{11} = 1 \geq 0.9 \\
\quad \quad - 2 \text{ consequences} \\
\quad \quad \cdot \{\text{risk=low} \Rightarrow \{\text{collateral=none, income=>35}\} \\
\quad \quad \quad \text{s(collateral=none, income=>35, risk=low)} = \frac{7}{14} = \frac{3}{9} < 0.9 \\
\quad \quad \cdot \{\text{income=>35} \Rightarrow \{\text{collateral=none, risk=low}\} \\
\quad \quad \quad \text{s(collateral=none, income=>35, risk=low)} = \frac{1}{7} = \frac{3}{8} < 0.9 \\
\quad \quad \cdot \{\text{collateral=none} \Rightarrow \{\text{income=>35, risk=low}\} \\
\quad \quad \quad \text{s(collateral=none, income=>35, risk=low)} = \frac{3}{11} = \frac{4}{11} < 0.9 \\
\end{align*}
\]
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 ≥90%). Note that these were verified with WEKA. WEKA seemed to have produced these except in a different order.

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