Data Mining
(Section 19.2)

Peter Revesz

CSCE 413/813
Computer Science and Engineering
University of Nebraska – Lincoln
Data Mining Example

A headhunting company needs to find an artistic and financially reliable applicant for director of an art foundation.

<table>
<thead>
<tr>
<th>Applicant</th>
<th>Name</th>
<th>Exhibitions</th>
<th>Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>70</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>80</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>10</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

Exhibitions = number of exhibitions made by the applicant. Debt = amount of personal debt in thousands of dollars.
Based on previous experience, a decision tree or a SVM can be used to identify a region of the feature space, which describes artistic and reliable.
The feature space regions can be represented by a constraint database.

### Personality

<table>
<thead>
<tr>
<th>Type</th>
<th>Exhibitions</th>
<th>Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>x</td>
<td>y ( t = \text{“Artistic”}, x \geq 40, x \leq 90, y \geq 10, y \leq 60, y \geq -x + 80 )</td>
</tr>
<tr>
<td>t</td>
<td>x</td>
<td>y ( t = \text{“Reliable”}, x \leq 80, y \geq -20, y \leq 20, y \leq x, y \geq -x )</td>
</tr>
</tbody>
</table>
The artistic and reliable applicants can be selected by a SQL query:

```
CREATE VIEW Promising_Candidate(Name)
SELECT Name
FROM Applicant AS A, Personality AS P
WHERE A.Exhibition = P.Exhibition  AND
A.Debt = P.Debt    AND
P.Type = "Artistic"
INTERSECT
SELECT Name
FROM Applicant AS A, Personality AS P
WHERE A.Exhibition = P.Exhibition  AND
A.Debt = P.Debt    AND
P.Type = "Reliable"
```
Data Mining

Challenge Question: List all the data mining tasks that the headhunting company is paid for.
Data Mining

Challenge Question: List all the data mining tasks that the headhunting company is paid for.

A. Creating a relational database for the job applicants.
B. Doing previous studies using decision trees or SVMs.
C. The previous studies likely included data integration.
D. Creating a constraint database for the relevant feature space regions.
E. Writing SQL queries for the identification of promising candidates.
F. Potentially also decision support in further ranking the candidates.