

ISO/IEC JTC 1/SC 32 N 0527

Date: 2000-09-12

REPLACES: --

<p>ISO/IEC JTC 1/SC 32</p> <p>Data Management and Interchange</p> <p>Secretariat: United States of America (ANSI) Administered by Pacific Northwest National Laboratory on behalf of ANSI</p>

DOCUMENT TYPE	Other document (Open)
TITLE	SQL/MM Data Mining Presentation
SOURCE	Friedmann Schwenkreis
PROJECT NUMBER	1.32.04.01.06.00
STATUS	This presentation will be given by WG 4 at the SC 32 Tutorials on 2000-10-9 in Helsinki
REFERENCES	
ACTION ID.	FYI
REQUESTED ACTION	
DUE DATE	
Number of Pages	7
LANGUAGE USED	English
DISTRIBUTION	P & L Members SC Chair WG Conveners and Secretaries

Douglas Mann, Secretariat, ISO/IEC JTC 1/SC 32

Pacific Northwest National Laboratory *, 901 D Street, SW., Suite 900, Washington, DC, 20024-2115,
United States of America

Telephone: +1 703 575 2114; Facsimile: +1 703 681 9180; E-mail: MannD@battelle.org

available from the JTC 1/SC 32 WebSite <http://www.jtc1sc32.org/>

*Pacific Northwest National Laboratory (PNL) administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI

SQL/MM Part 6: Data Mining

Friedemann Schwenkreis

2000-09-12

ISO/IEC 13249-6

1

What is data mining

- Find interesting patterns in the data
- Association rules
- Clustering/Segmentation
- Classification
- Regression
- ...

2000-09-12

ISO/IEC 13249-6

2

An association rule is a rule like:

If customer X buys product A the X also buys product B in 50% of all cases.

In 60% of all cases where A is bought, B is also bought.

Application: storage layout

Clustering:

Find segments with common characteristics. Find the number of segments and a prioritized list of attribute values which characterize the segments.

Application: target mailings

Classification:

Train a model with records for which we know the class such that we can use the model to predict the class for records without a given class. Minimize the number of attributes necessary to predict the class with an acceptable error.

Application: fraud detection

Regression:

Same as classification but the type of the predicted value. In case of regression it is a numeric value (e.g. number of insurance claims) rather than a class.

Application: cost prediction

This list can be extended.

Objectives

- **Bring data and data mining together**
- **Provide SQL interface to:**
 - Compute data mining models
 - Test data mining models
 - Apply data mining models
- **Allow data mining as part of SQL queries**

Out of Scope

- **Format of data mining models.**
- **Format of data mining metadata.**
- **Specific algorithms to implement a data mining technique.**

Future

- More mining techniques.
- More detailed functionality.
- Integration with statistics.
- Visualization support?
- Alignment with new parts of 9075 (e.g. SQL/XML).

Contact

Friedemann Schwenkreis
Editor of SQL/MM Part 6: Data Mining
fschwenk@de.ibm.com

2000-09-12

ISO/IEC 13249-6

6