

# *WizRule*®

Data mining tool for auditors

# Data Mining

*What is data mining?*

Data mining programs reveal interesting patterns in data.

**WizSoft** developed its data mining software tools based on association rules implementing a proprietary algorithm for revealing **if-then** rules in a given data.

The algorithm quickly reveals **all** the if-then rules that meet the user defined threshold.

# Auditing with data mining

- When the target is to underline suspected frauds without having previous examples, the auditor will use data mining.
- The data mining program reveals strong patterns in the data (the norm), then it points out at cases deviating from these patterns (anomalies).

# Auditing with data mining

- Data mining tools for auditing do not replace the standard audit tools, but rather complement them.
- The User doesn't tell the data mining program what to look for. The program does the 'thinking'.
- It works like a 'hounds', while the standard audit tools are more like 'watchdogs'.

# This is how *WizRule* works:

*WizRule* automatically reveals the main patterns (rules) that hold in the data and points out at cases deviating from these patterns as anomalies (suspected errors) or frauds.

This method does not detect cases when a fraud is done repeatedly because such a fraud creates a rule.

When a fraud is done repeatedly it creates a rule that deviates from other rules, and as such it is classified by *WizRule* as an outstanding rule.

# *WizRule*

analyzes the data by revealing:

- **If-then rules**
- **Formula rules** (among numeric fields)
- **Outstanding (if-then) rules**
- **Spelling rules**
- **If-then formula rules**

# Example of If-Then rule

*If*

***SUPPLIER\_NUMBER is J4175***

*Then*

***DEPARTMENT is 101.00 ... 400.00 (average = 337.63 )***

*Rule's probability: 1.000*

*The rule exists in 2326 records.*

*Significance Level: Error probability is almost 0*



# Example of If-Then rule

*If*

***NET\_INVOICE\_AMT is 15,073.44 ... 36,458.31 (average = 23,340.05)***

***and PAYMENT\_DATE is 4/1/2010 ... 8/11/2010 (average = 6/8/2010)***

*Then*

***INVOICE\_NUMBER is 4,999.00 ... 97,008,491.00 (average = 501,008.00)***

*Rule's probability: 1.000*

*The rule exists in 2,136 records*

*Significance Level: Error probability < 0.1*

**Rule Probability** - indicates the percentage of instances that both 'IF' and 'THEN' sections hold within the total number of cases in which 'IF' section holds – *Confidence Level*.

**Rule Exists in** - indicates the number of records in which the rule exists (IF and THEN applies) – *Support Level*.

**Significance Level** - indicates the probability that the rule exists non-accidentally.

# Example of Formula rule

$$A = B * C$$

*Where:*

**A = Total**

**B = Quantity**

**C = Unit Price**

*Rule's Accuracy Level: 0.99*

*The rule exists in 1890 records*

**Accuracy Level** indicates the ratio between the number of cases in which the formula holds and the total of cases of the rule itself

# Example of Outstanding rule

*If*

**Customer is Summit**

*Then*

**% Discount = 20**

*The rule is unexpected since:*

There are **100** values in the **Customer** field, each having no less than the minimum number of cases in a rule. Found no similar rules that relate between the other values in the **Customer** field and the values in the **%Discount** field.

# When is a rule **Outstanding**?

- a) It is the only rule that relates between a certain customer and a certain % discount (while all the customers have various discounts).
- b) If the % discount of all the other customers were 10% for example, the above mentioned rule would still be outstanding, since the discount of this customer deviates from the discount of all the other customers.

# A rule is **Outstanding**:

- a) When it is the only rule that relates between specific values of two fields. No similar rules that relate between other values in both fields were found.
  
- b) When it is unexpected relatively to other rules representing the norm (in other words, another rule is expected).

# Example of **Spelling** rule

*The value **Summit** appears **2080** times in the **Customer** field.*

*There are **2** case(s) containing similar value(s)*

The spelling rules are presented mainly in order to reveal cases of misspelled names. A name is suspected as misspelled if (a) it is similar to another name in this field, and (b) the frequency of the first name is very low, while the frequency of the second name is very high.

# Example of If-Then Formula rule

*If*

Customer is Summit

*Then*

**Total = Unit Price \* 0.8 + 50**

*Accuracy Level: 0.95*

*The rule exists in 502 records*

**Accuracy Level** indicates the rate of cases that fit the formula (plus or minus a pre-determined limits) relative to the total number of cases where the condition holds.



# *WizRule*'s reports:

**Overview Report** - displays the value distribution in each field of the dataset, providing data insights.

**Rule Report** - lists all the discovered rules.

**Spelling Report** - lists cases that are likely to be spelling errors of names or other values.

**Deviation Report** - lists the cases that deviate from the set of the discovered rules (cases to be audited).

*Note: a **deviation** from a highly valid rule might point to a fraud.*

**Outstanding Rules Report** - lists unexpected rules to view interesting phenomena that may lead to a quick discovery of fraudulent cases.

# WizSoft®

**516-393-5841**

**[info@wizsoft.com](mailto:info@wizsoft.com)**