

WizWhy[®]

Analyzer & Predictor

Data Mining

What is data mining?

Data mining programs reveal interesting patterns in data.

WizSoft developed its predictive analytics tool (*WizWhy*) based on association rules, implementing a proprietary algorithm revealing **all** the ***if-then*** rules that explain one field by other fields.

The ***if-and-only-if*** rules are revealed and they represent the *necessary and sufficient* conditions. These rules are optimal, as they cover the maximum number of both positive and negative cases.

WizSoft holds a patent for SYSTEM AND METHOD FOR REVEALING NECESSARY AND SUFFICIENT CONDITIONS FOR DATABASES ANALYSIS – Patent No: US 6,542,881 B1, Apr. 1, 2003.



WizWhy

analyzer & predictor

- Easy to learn
- Easy to use
- Simple results to understand and explain

WizWhy

analyzer & predictor

WizWhy determines how the values of one field (*dependent variable*) are affected by the values of other fields (*independent variables*).

WizWhy performs the analysis for one field, selected by the user as the *dependent variable*, while the other fields are the *independent variables* (conditions).

The analysis can be either Boolean or multi-value (categorical variable) or continuous (numeric variable).

WizWhy

analyzer & predictor

WizWhy learns the patterns of positive and negative examples of the dependent variable and issues predictions for new cases.

When issuing predictions *WizWhy* applies the discovered rules (*the model*) to another dataset (from the same population) and predicts the outcome for each case (record).

This is how *WizWhy* works:

WizWhy first *reads* the data. The user *selects* the dependent variable, *fine-tunes* the analysis by defining the minimum probability of the rules, the minimum number of cases in each rule, and the cost of a miss vs. the cost of a false alarm.

WizWhy *analyzes* the data and *issues* the rules that relate between the dependent variable and the other fields (conditions). The rules are formulated as '*if-then*' and '*if-and-only-if*' sentences.

This is how *WizWhy* works:

On the basis of the discovered rules *WizWhy*:

- **summarizes** the data
- points out **interesting phenomena**
- reveals **necessary & sufficient conditions** (main patterns)
- points out **unexpected cases** (cases to be audited)
- issues **predictions** for new cases

The predictions can be:

- **Boolean** (e.g., whether the company will go bankrupt or not)
- **continuous** (e.g., the %Growth is between 5% - 7%)
- **multi-value** (e.g., given the patient's symptoms, what the disease is)

WizWhy

analyzes the data by revealing:

- all the **if-then** and **if-then-not** rules
- **if-and-only-if** rules (*necessary & sufficient conditions*)
- **if-then formula** rules (when the dependent variable is numeric and continuous)

Example of **if-then** rule

If

City is New-York

and **Amount Purchased is 200 ... 300 (average = 250)**

and **Sale Person is Dave**

Then

Growth since Last Year is less than 1%

Rule's probability: 0.70

The rule exists in 570 records

Significance Level: Error probability < 0.001

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*'.

Error probability - indicates the degree to which the rule can be relied upon as a basis for predictions. Assuming that the data under analysis is a representative sample of an infinite population, the error probability quantifies the chances that the rule does not hold in the entire population and exists accidentally in the file under analysis.

Example of **if-and-only-if** rule

The following conditions explain when the Growth since Last Year is less than 1%:

The conditions are:

- (1) The Amount Purchased is between 0 ... 199 (average = 100)**
- (2) The Sale Person is Dan and the City is Boston**

If at least one of these conditions holds, the probability that the Growth since Last Year is less than 1% is 0.9

If none of these conditions holds, the probability that the Growth since Last Year is *not* less than 1% is 0.95

i.e., the Growth since Last Year is less than 1%, **if-and-only-if**, the Amount Purchased is between 0 ... 199 **or** the Sale Person is Dan and the City is Boston.

If-then rules represent sufficient conditions.

The **if** condition is a sufficient condition for the result (**then** part).

If-and-only-if rules go one step further; they represent necessary and sufficient conditions.

In the previous example, the two conditions, (1) and (2), are necessary and sufficient conditions for the Growth being less than 1%.

Revealing the **if-and-only-if** rules:

- ✓ points out the main patterns in data.
- ✓ solves the problem of too numerous **if-then** rules that practically cannot be read in order to understand the data.

Since **WizWhy** searches for the optimal **if-and-only-if** rules, each value of the dependant variable is explained by one or two **if-and-only-if** rules.

The **if-and-only-if** rules are optimal in the sense that they cover the maximum number of both positive and negative examples.

Example of **if-then formula** rule

If

City is New-York

Then

$$\mathbf{A = B * 0.4 + 35}$$

where: A = Growth Since Last Year

 B = Amount Purchased

Accuracy level: 0.95

The rule exists in 170 records

Accuracy Level indicates the percentage of records in which the formula holds (within the formula limits determined when issuing the rule report) relative to the records regarding which the condition (the if clause) holds.

WizWhy issues the following reports:

Summary Report -

1. Lists the parameters of the analyzed data together with the user's parameters, the parameters according to which the rules were issued.
2. Displays an analysis of the rules explanatory power, to what extent the discovered rules explain the data; as well as displays the results which set of rules explain the dependent variable most accurately - if-then rules or if-and-only-if rules.

Rule Report - lists the discovered **if-then** and **if-then-not** rules.

WizWhy issues the following reports:

Trend Report – presents graphically and textually the one condition trends in the data. *WizWhy* **summarizes** the data by listing the relations between each field and the dependent variable, and calculates the explanatory power of each field.

Unexpected Rules Report – lists the rules that are **unexpected** relative to basic rules and trends. These unexpected rules describe interesting phenomena in the data.

WizWhy issues the following reports:

If-and-only-if Rules Report – lists the *necessary and sufficient* conditions.

Unexpected Cases Report – lists cases (records) where the dependent variable's value deviates from the *expected value* according to the discovered rules. These are cases deviating from the rules, cases to be audited.

WizWhy issues predictions

When the analysis is completed, the model is created and *WizWhy* can issue predictions for new cases.

WizWhy predicts on the basis of the discovered rules.

The data of the new cases to be predicted may be entered either manually or read through a given file.

WizWhy calculates the value of the prediction and its *conclusive probability*. These figures indicate the extent to which the 'subject to predict' can be relied on. The predictions can be either Boolean or continuous or multi-value.

WizWhy issues predictions

Prediction methods:

- **Predict to File:** the input for issuing predictions comes from a given file and the predictions results are saved to another file.
- **Predict On-Line:** the condition fields are entered manually and the prediction results are displayed on the screen.
- ***WizWhy* Predictor:** it's an independent application for issuing predictions on the basis of the discovered rules by *WizWhy*. This method is ideal for limiting access to the contents of confidential data.

WizSoft®

516-393-5841

info@wizsoft.com