

## **Dr. Ferenc Kiss: Evolution and Application of Credit Scoring**

**Thesis1.** I have developed a new taxonomy and a classification model for judging credit scoring methods in knowledge management point of view.

Knowing the theoretical and methodological background of the credit scoring models and the modelling methods it is statable some of these procedures fits schemes to experienced, historical data, while another part of them convert the experiences and knowledge of experts and the goals and policies of the credit grantors into rules for building the models. So a new classification framework required which is heavily different than the methodological based one. The categories as follows:

- *Knowledge generation credit scoring modelling methods.* All the methods which generate the decision model using statistical and another analytical techniques based on historical database. This means a kind of (automatic or manual) formalization of the hidden experiences in the data.
- *Knowledge preservation credit scoring modelling methods.* These procedures are converting and formalizing the experiences and knowledge of experts. Thus this collected knowledge will be available, reachable continuously.
- *Knowledge selection credit scoring modelling methods.* The members of this group are the procedures which are suitable for building decision models for selecting the nearest optimal model from the available set of existing models according to the gives decision situation.

According to these categories the classification of credit scoring methods in knowledge management point of view is the following, based on the main characteristics of the procedures.

#### Knowledge generation credit scoring modelling methods

- Linear probability models;
- Probit and Logit models;
- Discriminant analysis based models;
- Neural networks;
- Mathematical programming;
- Classification trees (recursive partitioning algorithms);
- k-Nearest Neighbours.

#### Knowledge preservation credit scoring modelling methods

- Analytical Hierarchy Process;
- Expert systems.

#### Knowledge selection credit scoring modelling methods

- Decision trees;
- Genetic algorithms.

Because this classification is one-sided regarding some of the methods, it is suitable to declare these three categories to perpendicular dimensions to a classification space. Only this model is capable to describe the attributes of each method exactly. That's why the description of the credit scoring method in the knowledge management point of view is not trivial.

However using this new classification system offers a standardized way of judging for all the existing and the future new methods.

**Thesis2.** I prepared a new evaluation scale for judging the available set of databases for building credit scoring models.

The question whether the new credit scoring model should be developed for an existing product or product chain or may be for a very new product is fundamental for the whole development process. This mostly determines what kind of historical databases, business experiences will be available at the starting situation.

If there is a historical database relating to the targeted customers, its content of information, representativeness, the volume of missing

values, the data cleanse and quality are key aspects for the usability in the model development process and the probable effectiveness and accuracy of the produces model.

Based on these attributes a new ordinal scale could be defined for measuring the usability of the available databases. The D1..D11 categories of this dimension are summarized in the next table.

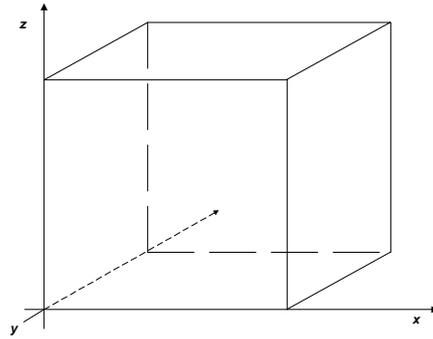
D1	No any relating dataset or database
D2	<u>Only general statistical data</u> are available about the targeted customers (e.g. National Statistical Office database, census data, demographical, income, etc. distributions)
D3	Besides the general statistics there are <u>descriptive statistics</u> coming from own other type business transaction history or bought <u>directly about the targeted customers</u> .
D4	There <u>is a historical database coming from similar business transaction history</u> , with varying confidence and missing values.
D5	There <u>is a representative historical database</u> coming from similar business transaction history (HDB), with varying confidence and missing values.
D6	There is an HDB, there is <u>no missing value among the primary data</u> , the rest fields are with <u>varying confidence and less than 20% missing values</u> .
D7	There is an HDB, there is <u>no missing value among the primary data</u> , the rest fields are with <u>confidence above 80% and less than 20% missing values</u> .
D8	There is an HDB, there is no missing value among the primary data, the rest fields are with confidence above 80% and <u>less than 5% missing values</u> .
D9	There is an HDB, there is no missing value among the primary data, the rest fields are with <u>confidence above 95%</u> and less than 5% missing values.
D10	There is an HDB, with <u>no missing value at all</u> and with confidence above 95%.
D11	There is an HDB, with <u>no missing value at all</u> and with 100% confidence.

**Thesis3.** I prepared a new decision model for selecting a credit scoring modelling method depending on the given business situation.

Business (product) experience and the knowledge of the targeted customers are required for building credit scoring models, especially if there are little experience (transactional) data available. Both expertise might be available in the organization or from outside of it. According to these an ordinal scale dimension could be defined for measuring the competence, as the next table describes.

E1	There is no business experience on the given decision situation.
E2	There is only basic, general knowledge about the given decision situation.
E3	There are little experience and knowledge about the product and the targeted customers as well.
E4	There are some years experience and knowledge about the product and the targeted customers as well, but the product is new, or almost new that means there are few closed transaction.
E5	There is a wide scale of long term experiences and knowledge about the product and the targeted customers as well.

During the evaluation of a given decision situation these two dimensions should be applied. In this case it is determinable which method is worth to select for building a credit scoring system.



The classification space of the probable effectiveness of credit scoring methods

where: - x axis shows the evaluation of the database,  
 - y axis shows the evaluation of the competence,  
 - z axis shows the probable effectiveness of each methods.

The probable effectiveness of credit scoring methods could be evaluated on the following ordinal scale:

H1	Usage of the method probably will be unsuccessful
H2	Usage of the method probably provides a model with major unreliability
H3	Usage of the method probably provides a model with acceptable accuracy
H4	Usage of the method probably provides a model with appropriate accuracy