Using Data Mining Approaches to Build Credit Scoring Model

Abstract

- The aim of this paper is to present a Credit Scoring Model applied by one Microfinance Institution in Bosnia and Herzegovina as well as to show how the most relevant attributes for its implementation were identified.

- The purpose of created Credit Scoring Model was to predict default clients and reduce credit risk of Microfinance Institution by applying data mining algorithm in order to find patterns for recognition of default clients and, thus, support decision making process of credit approval.

- Credit Scoring Model was build using Oracle Data Miner software package that uses Generalized Linear Model for classification.

- Created model showed great predictive confidence and accuracy, but also gave trustworthy results regarding feature selection, so the Microfinance institution decided to adopt this model as help in decision making process.

Existing

- Credit Scoring is used very often as the main tool for risk management in financing institutions.

- This model represents set of decision models built on top of advanced data mining techniques and algorithms that involve extraction of hidden patterns and information, and are used to predict default clients which is one of the crucial factors when approving credit loan requests.

- The need for implementation of credit scoring emerged as a consequence of bad decisions in loan approval that relied on previous staff experience and subjective

disadvantage

- Need description to how Data Mining algorithms were applied to select the most important attributes for implementation of an efficient and highly reliable Credit Scoring Model which would be used as a tool of support in decision making process when approving client's loan request.

- Financial institutions are faced with many dilemmas regarding loan approving process, because apart of taking risk of approving loan to the client who can become defaulter, they are also missing chance to make profit by declining loan to the client who is capable of paying his obligations. This is why many financial institutions lately are adopting credit scoring models that are capable to recognize hidden patterns in very large databases in order to classify clients as default or non-default

Proposed

- the process of applying data mining algorithms in order to identify and select the most relevant attributes from the given dataset and build accurate and efficient credit scoring model.

- This study is a part of a real life project of Credit Scoring Model appliance in loan request approval process, in a finance institution.

- Thus, it shows in detail how advanced data mining algorithms helped this company improve and optimize their business processes.

Advantage

- this model is very adjustable, and thus, it can be applied to other datasets and other Microfinance institutions interested into optimization of their decision making processes

- This will not only help our Microfinance institution to accelerate their business processes but also to manage their risks better and reduce costs as well

- used for support of decision making process when approving loan request and, thus, predicting default clients

HARDWARE REQUIREMENTS

- Processor   -   Pentium –III
- Speed   -   1.1 Ghz
- RAM   -   256 MB(min)
- Hard Disk   -   20 GB
- Floppy Drive   -   1.44 MB
- Key Board   -   Standard Windows Keyboard
- Mouse   -   Two or Three Button Mouse
- Monitor   -   SVGA

MICANS INFOTECH

SOFTWARE REQUIREMENTS

- Operating System　　　　：　Windows 8
- Front End　　　　　　　：　Java /DOTNET
- Database　　　　　　　：　Mysql/HEIDISQL

MICANS INFOTECH

Conclusion

▸ This paper is based on the real project whose main goal was to provide managing of credit risk to financial institutions.

▸ Financial institutions are faced with many dilemmas regarding loan approving process, because apart of taking risk of approving loan to the client who can become defaulter, they are also missing chance to make profit by declining loan to the client who is capable of paying his obligations.

▸ This is why many financial institutions lately are adopting credit scoring models that are capable to recognize hidden patterns in very large databases in order to classify clients as default or non-default.

▸ Our research includes appliance of data pre-processing and feature selection techniques on real dataset provided by a Microfinance institution in order to build accurate and efficient Credit Scoring Model with short running time based on data mining algorithm for classification, logistic regression.

▸ Created Credit Scoring Model showed unexpectedly good accuracy and recognize a 100% of default clients. This model was validated by representatives of the institution that will be used for support of decision making process when approving loan request and, thus, predicting default clients.

▸ This will not only help our Microfinance institution to accelerate their business processes but also to manage their risks better and reduce costs as well.

▸ Finally, we have proved that this model is very adjustable, and thus, it can be applied to other datasets and other Microfinance institutions interested into optimization of their decision making processes.

Future work

- Since we achieved high level of accuracy with our credit scoring model, and that it gained trust of Microfinance institution as a reliable tool, we agreed that they will tend to input even more accurate and detailed data even for the attributes that we had to fill in with unknown values during pre-processing phase.

- This will enable us to rebuild model again in near future, and develop new, improved model with the same or even higher accuracy.

Reference

[1] L. Abid, A. Masmoudi and S. Zouari-Ghorbel, "THE CONSUMER LOAN'S PAYMENT DEFAULT PREDICTIVE MODEL: AN APPLICATION IN A TUNISIAN COMMERCIAL BANK," Asian Economic and Financial Review, vol. 6, pp. 27-42, January 2016.

[2] S-C. Huang and M-Y. Day, "A Comparative Study of Data Mining Techniques for Credit Scoring in Banking", IEE IRI San Francisco, California, USA, August 2013.

[3] M. P. Bach, J. Zoroja, B. Jakovićand N. Šarlija, "Selection of Variables for Credit Risk Data Mining Models: Preliminary Research", MIPRO Opatija, Croatia, May 2017.

[4] L. Feng-Chia, W. Peng-Kai and Y. Li-Lon, "Diversity of Feature Selection Approaches Combined with Distinct Classifiers", 2010 IEEE International Conference on Industrial Engineering and Engineering Management, December 2010.

[5] E. Kocenda and M. Vojtek, "Default Predictors in Retail Credit Scoring: Evidence from Czech Banking Data", William Davidson Institute Working Paper Number 1015, April 2011.