SECOND INTERNATIONAL CONFERENCE
Modelling and Development of Intelligent Systems
Sibiu - Romania, September 29 - October 02, 2011

The Analysis of Continuous Variables in the Decision Model of Bankruptcy Risk using Bayesian Networks

Mihaela Crăciun, Dominic Bucerzan, Crina Rațiu

Abstract

This paper is concerned with the modeling using bayesian network (BN) of bancruptcy prediction (BP) from the economic model proposed by Anghel [5]. Within the simulation the paper is focused on the choosing a discretizing method for the used interval, depending on the performance of the three methods chosen. Comparison is made between Bracket Medians discretizing method and Pearson-Tukey discretizing method. The BN construction process, respectively the simulation was realized using the AgenaRisk software. Simulation results were obtained from sensitivity analysis table and graphic.
Keywords: *Bayesian Network (BN), Normal Distribution, Discretizing, Bracket Medians Method, Pearson-Tukey Method*

1 Introduction

Modeling using bayesian network (BN) of bancruptcy prediction (BP) from the economic model proposed by Anghel [5], is presented in this work.
For simulation we have chosen several discretizing methods for the used interval.
Before detailing the proposed solution, we define de necessary elements for modeling using BN.

Random variables

Given a probability space (Ω, P), a ***random variable*** X is a function whose domain is Ω. The range of X is called the space of X.
We call P(X=x) the ***probability distribution*** of the random variable X. [1]

Bayesian Network – BN

Bayesian networks – BN consist of:
- a direct acyclic graph (DAG), whose edges represent relationships among random variables that are often (but not always) causal;
- the prior probability distribution of every variable that is a root in the DAG; and
- the conditional probability distribution of every non-root variable given each set of values of its parents. [2]

The I. Anghel Model in Bankruptcy Risk Prediction – Anghel PM for BR

Anghel PM for BR is based on a statistical data sample collecting during the period 1994-1998. The author used the discriminant analysis method (MDA) for selecting bankrupt enterprises. Thus is created the following function score.
$A = 5.667 + 6.3718 * X_1 + 5.3932 * X_2 – 5.1427 * X_3 – 0.0105 * X_4$, subject is founded in [5].
Where:

X1 - earning after taxes / incomes;
X2 - Cash Flow / total assets;
X3 - liability / total assets;
X4 - liability/ sales * 360

Weight coefficients were calculated according to the variation and covariation of the financial variables and according to the difference of the variables mean broken down by enterprises groups: bankrupt and good situation.

The inflection point that minimize the error rate is A=0, with an uncertainly range between 0 and 2.05.

In case of *a prior* analysis of the success rate for the score function A, it will compared the predictive classification with the known situation of the enterprises in the sample. When choosing an inflection point equal to 0 it determine a success rate of 91.2%. In case we choose two inflection points (A=minim 0 and A= maximum 2.05) and we consider a zone of uncertainty between this two points, we obtain a success rate of 96.7%.

The risk assessment in Anghel PM for BP is the following:
When $A < 0$, bankruptcy/failure situation;
When $0 \le A \le 2.05$, uncertainty situation demanding prudence;
When $A > 2.05$, a good financial situation

In case of *a posterior* analysis of the success rate for the score function A, it will analyzed the degree of relevance for another sample of enterprises. It will obtain a prediction success rate without an uncertainly zone of 92.8%, respectively the success rate with an uncertainly zone of 97.8%.

The results obtained in the two analysis permit assessment that the A score is efficient and can be applied to enterprises in the Romanian economy.

Discretizing

Let be a BN that contains random variables that are discrete or continuous. For the continuous variable the possible values of the node are ranges and the probability of each of these ranges is specified in the network. This is called discretizing the continuous variables. [3]

Methods for discretizing

A. Bracket Medians Method [4]

In the Bracket Medians (BM) Method the mass in a continuous probability distribution function $F(x) = P(X \le x)$ is divided into n equally spaced intervals. The method proceeds as follows. Typically we can use three, four ore five intervals. If we have more intervals, the computation is more accurate. Let be n=3 in this explanation. Next we use the BM Method to discretize the distribution into three ranges. They are four steps to follow:

Step 1: For values between 0 and 100, we consider following three intervals:

$$[0, 0.333], [0.333, 0.666] \text{ and } [0.666, 1].$$

Step2: We need to find points x_1, x_2, x_3 and x_4 such that:

$$P(X \le x_1) = 0.0,\ P(X \le x_2) = 0.333,\ P(X \le x_3) = 0.666,\ P(X \le x_3) = 1$$

where the values on the right in these equalities are the endpoints of the three intervals.

Using mathematics package we obtain: x_2=43.5 and x_3=56.4. Clearly we have x_1=0 and x_4=100.

Step 3: For each interval $[x_i,x_{i+1}]$ we compute the bracket median d_i, which is the value such that: $P(x_i \le X \le d_i)=P(d_i \le X \le x_{i+1})$.

Step 4: Define the discrete variable D with the following probabilities:

$$P (D=d_1) = 0.333,\ P (D=d_2) = 0.333,\ P (D=d_3) = 0.333$$

B. Pearson-Tukey Method [4]

In the Pearson-Tukey Method the mass in a continuous probability distribution function $F(x) = P(X \le x)$ is divided into three intervals. The method proceeds as follows:

Step 1: Determine points x_1, x_2 and x_3 such that

$$P(X \le x_1) = 0.05,\ P(X \le x_2) = 0.50,\ P(X \le x_3) = 0.95$$

Step2: Define the discrete variable D with the following probabilities:

$$P (D = x_1) = 0.185,\ P (D = x_2) = 0.63,\ P (D = x_3) = 0.185$$

Using mathematics package we obtain the cut points 36.6 respectively 63.4.

2 Bayesian Networks Construction

We explain the Anghel PM for BR accepting the Bayes' Theorem and the accuracy of the software AgenaRisk [6] in [7].

After the BN construction we obtain the visual model as shown in the Figure 1.

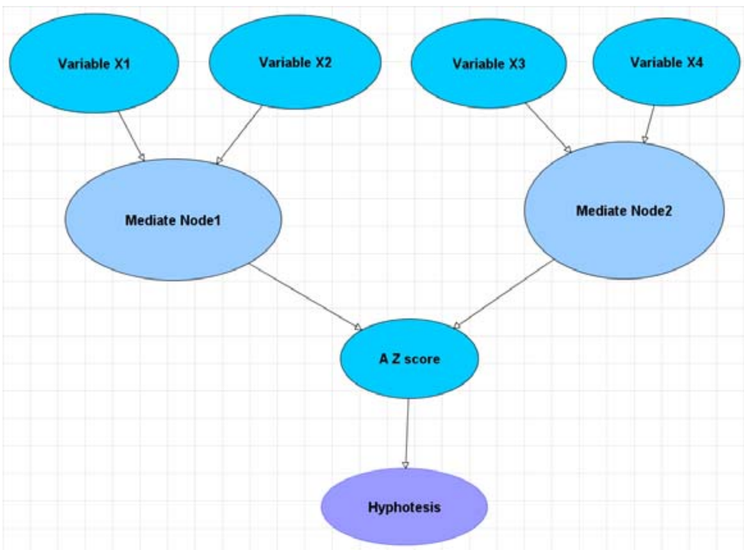


Figure 1 – BN showing causal structure

We use three types of nodes to model our BN: sample, result and assumption nodes.

The difference between the BN presented in [7] and the model presented in this paper refers to the lower and upper bounds we consider and the graph type we associate. Also we removed from de model the observation nodes.

The *sample nodes* are simulation nodes, with continuous interval type. The lower bound is 0 and the upper bound is 100. The NPT is a Normal Expression with mean 50 and variance 225. The graph types associated to this node are Line and represent the Probability Distribution.

The *result nodes* are simulation nodes, too. They divide in two categories. The Mediate Nodes and the A Z score node. The types of Mediate Nodes are continuous interval with values between 0 and 1.176. The NPT is an arithmetic expression $6.3718*X_1 +5.3932* X_2$, respectively $5.667-5.1427*X_3-0.0105*X_4$. The graph types associated to this node are Line and represent the Probability Distribution. The type of A Z score node is continuous interval with values between -520 and 2.700. The NPT is an arithmetical expression MN1+MN2. The graph type associated to this node is Line and represents the Probability Distribution.

The *assumption node* Hypothesis is a simulation node, with Boolean type. The state options are customised, with Positive Outcome “Good financial situation” and the Negative Outcome “Bankruptcy / failure situation”. The NPT is a comparison expression: ***if(zscore<=2.05,"Bankruptcy / failure situation", "Good financial situation")***. The graph type associated to this node is Line and represents the Probability Distribution.

The statistic attached to the main risk graph is shown in Figure 2.

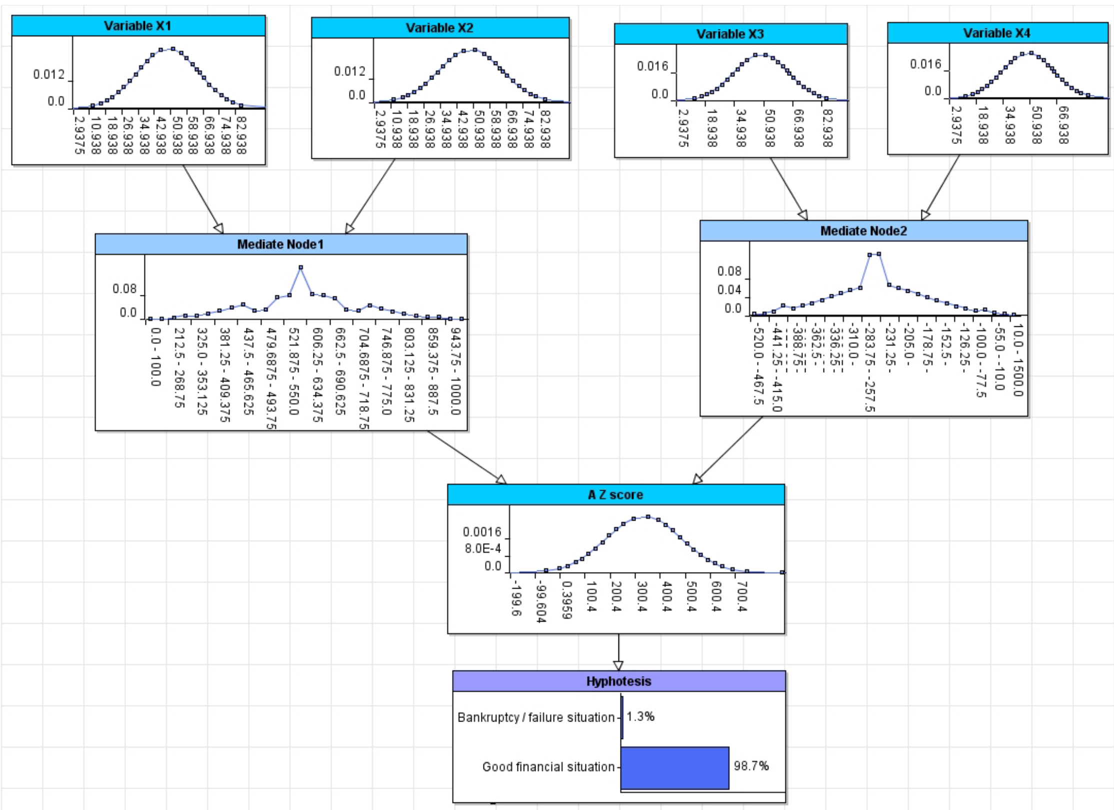


Figure 2 - Complete Hypothesis Testing Model

3 Hyphotesis Testing Model Simulation

Within the *sample nodes* we define the Node Probability Table – NPT, using the Normal Distribution expression, see the Figure 3.

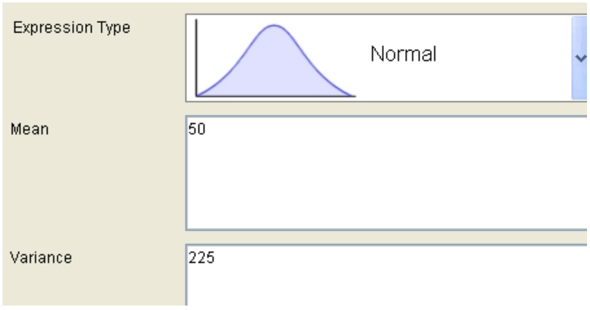


Figure 3 – The NPT defined as Normal Distribution with Mean=50 and Variance=225

3.1 Simulation using one interval

This simulation case is based on the BN structure defines by section 2. It works of the interval [0, 100], defined in the *sample nodes*. In this case we obtain a percentage of 98.700 % “Good financial situation” (see Figure 2). In simulation with scenario we test as enter observation in sample node Variable X1 the value 25 (mean between 20 and 30, see values obtained in Figure 11). In this case we obtain a percentage of 93.633 % “Good financial situation”.

After the Sensitivity Analysis we obtain the results shown in Figure 4.

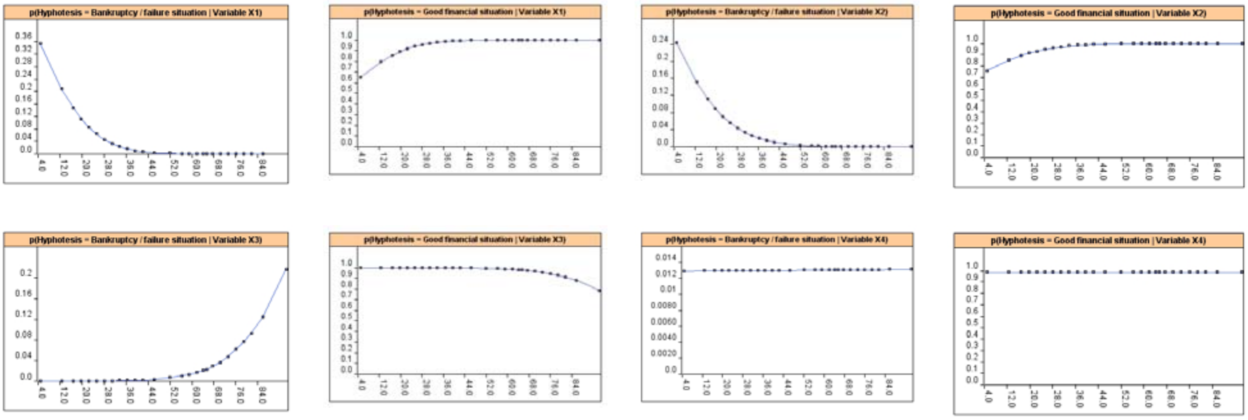


Figure 4 – Sensitivity Analysis in the interval [0, 100]

3.2 Simulation using Bracket Medians (BM) Discretizing Method

In this case, we must first build our BN. The key for this BN is defined in the *sample nodes*, more accurate in his Node States (see Figure 5) and in his Node Probability Table (see Figure 6).

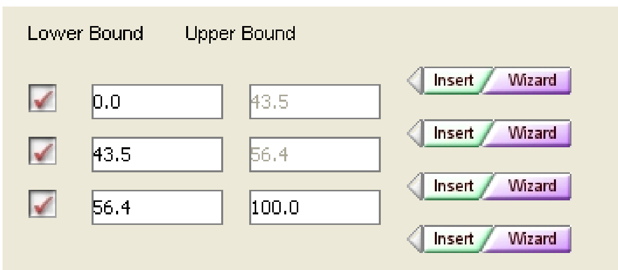
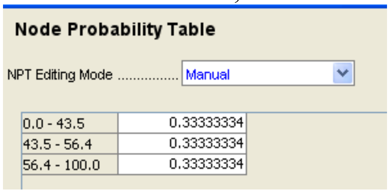


Figure 5 – Node States for the sample nodes in BM Discretizig Method

After defining all the nodes like in the section 2, we obtain the BN shown in the Figure 7.


Node Probability Table

NPT Editing Mode .... Manual

0.0 - 43.5	0.33333334
43.5 - 56.4	0.33333334
56.4 - 100.0	0.33333334

Figure 6 – NPT in BM Discretizing Method

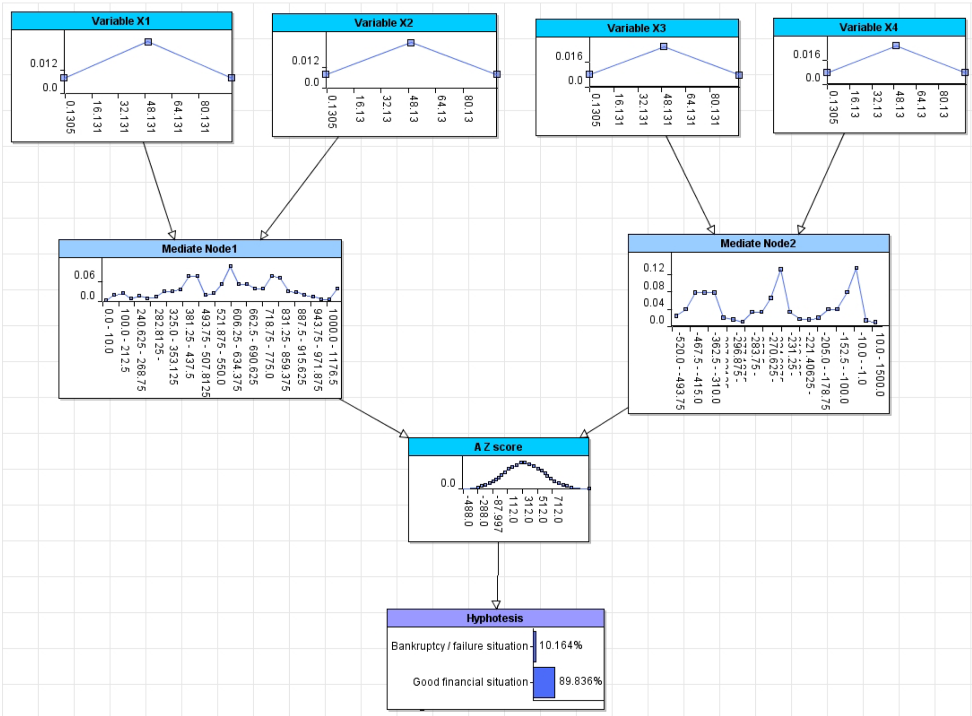


Figure 7 – Complete Hyphotesis Testing Model using BM Discretizing Method

In the BM Method we work within three intervals [0, 43.5], [43.5, 56.4] and [56.4, 100] defined in the *sample nodes*. In this case we obtain a percentage of 89.836 % "Good financial situation" (see Figure 7).

After the Sensitivity Analysis we obtain the results shown in Figure 8.

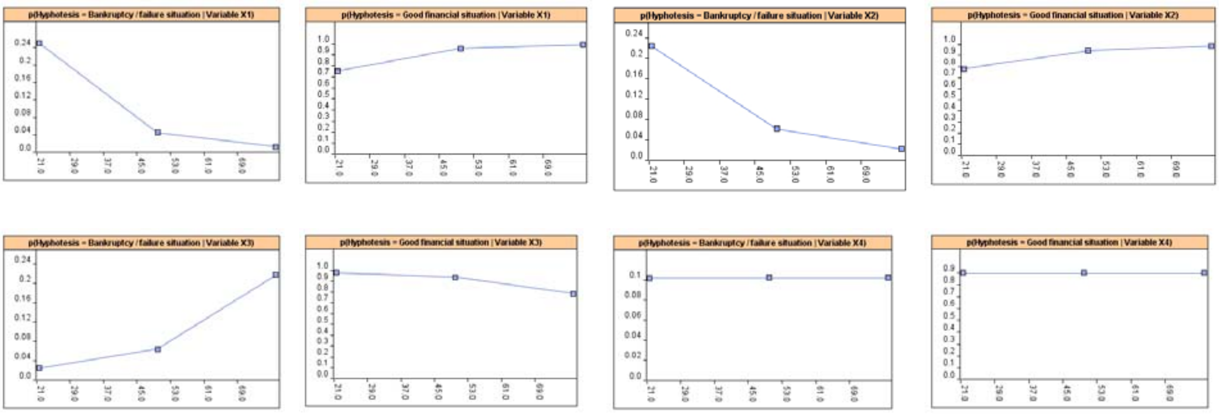


Figure 8 - Sensitivity Analysis using Bracket Medians Discretizing Method

3.3 Simulation using Pearson-Tukey (P-T) Discretizing Method

In this case, we must first build our BN, just like in the subsection 3.2. The key for this BN is defined in the *sample nodes*, more accurate in his Node States (see Figure 9) and in his Node Probability Table (see Figure 10).

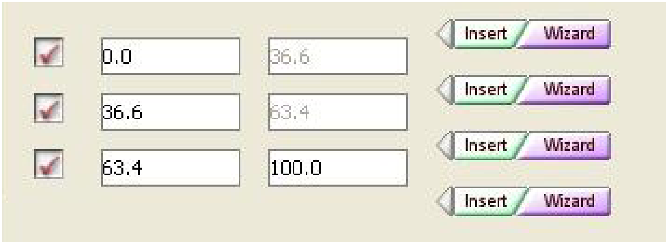
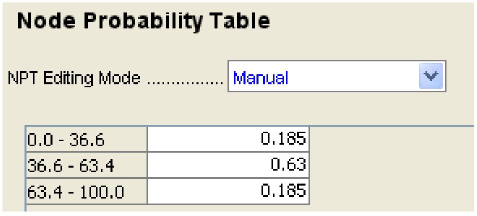


Figure 9 - Node States for the sample nodes in P-T Discretizing Method



0.0 - 36.6	0.185
36.6 - 63.4	0.63
63.4 - 100.0	0.185

Figure 10 – NPT in P-T Discretizing Method

After defining all the nodes like in the section 2, we obtain the BN shown in the Figure 11.

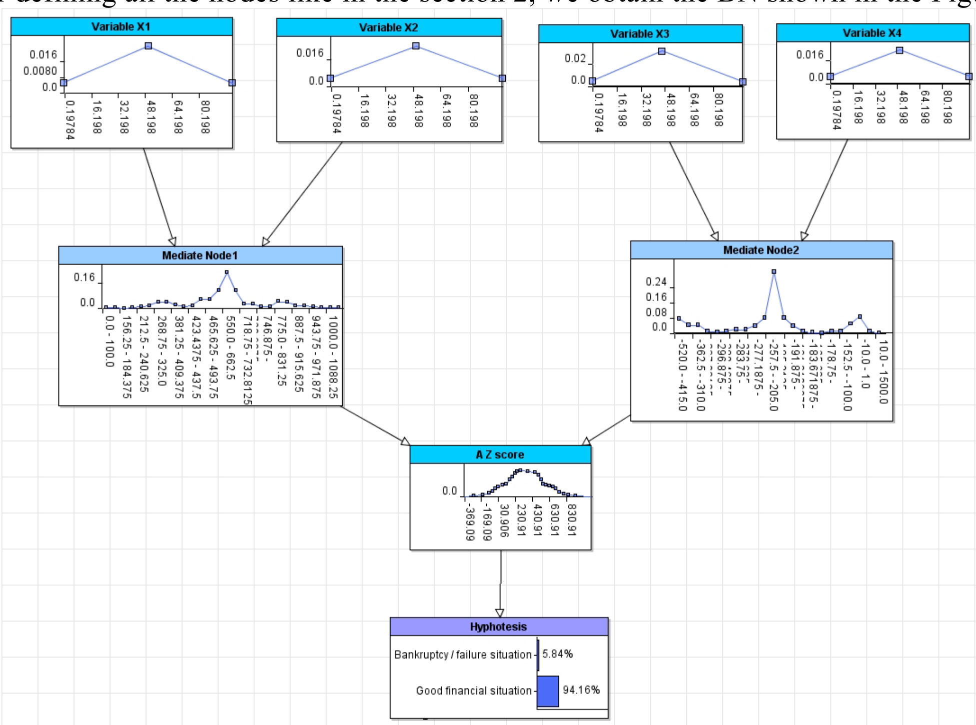


Figure 11 – Complete Hyphotesis Testing Model using P-T Discretizing Method

In the P-T Method we work within three intervals [0, 36.6], [36.6, 63.4] and [63.4, 100] defined in the *sample nodes*. In this case we obtain a percentage of 94.16 % “Good financial situation” (see Figure 11).

After the Sensitivity Analysis we obtain the results shown in Figure 12.

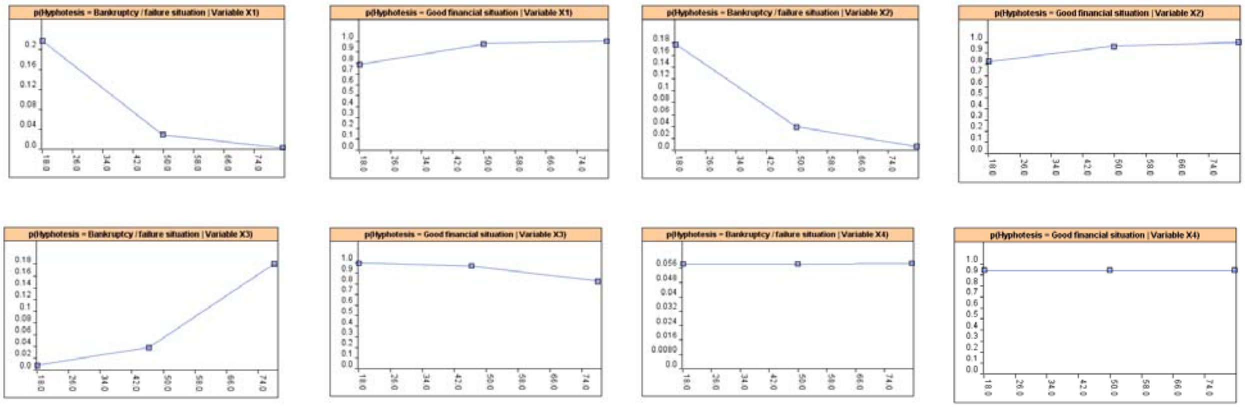


Figure 12 - Sensitivity Analysis using Pearson-Tukey Discretizing Method

4 Conclusions

The Hyphotesis Testing Model results obtained in our paper are highlighted in Table 1.

	1 Interval without scenario	1 Interval with scenario	Bracket Medians	Pearson-Tukey
Bankruptcy / failure situation	*1,300%*	*6,367%*	*10,164%*	*5,84%*
Good financial situation	*98,700%*	*93,633%*	*89,836%*	*84,160%*

Table 1 – Comparison between Hyphotesis Testing Model results using the presented methods

BM Discretizing Method Analysis

p(Hyphotesis | Variable X1)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X1	21,75	0,249	0,751
	49,95	0,045	0,955
	78,20	0,012	0,988
		0,306	

p(Hyphotesis | Variable X2)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X2	21,75	0,222	0,778
	49,95	0,061	0,939
	78,20	0,021	0,979
		0,304	

p(Hyphotesis | Variable X3)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X3	21,75	0,024	0,976
	49,95	0,064	0,936
	78,20	0,217	0,783
		0,305	

p(Hyphotesis | Variable X4)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X4	21,75	0,101	0,899
	49,95	0,102	0,898
	78,20	0,102	0,898
		0,305	

	Bankruptcy / failure situation
Sensitivity Analysis	30,50%
Hyphotesis in BM	10,16%
	20,34%

P-T Discretizing Method Analysis

p(Hyphotesis | Variable X1)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X1	18,3	0,216	0,784
	50,0	0,029	0,971
	81,7	0,002	0,998
		0,247	

p(Hyphotesis | Variable X2)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X2	18,3	0,177	0,823
	50,0	0,039	0,961
	81,7	0,006	0,994
		0,222	

p(Hyphotesis | Variable X3)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X3	18,3	0,007	0,993
	50,0	0,037	0,963
	81,7	0,181	0,819
		0,225	

p(Hyphotesis | Variable X4)

	Hyphotesis	Bankruptcy / failure situation	Good financial situation
X4	18,3	0,058	0,942
	50,0	0,058	0,942
	81,7	0,059	0,941
		0,175	

	Bankruptcy / failure situation
Sensitivity Analysis	21,73%
Hyphotesis in BM	5,84%
	15,89%

Figure 13 – Sensitivity Analysis in BM and P-T Discretizing Methods

The influence of sample nodes X_i prediction on the assumption node Hyphotesis prediction is analyzed in Figure 13.

Mean of the bankruptcy prediction for the sample nodes X_i in BM discretizing method is 30.50%. Also in this method the bankruptcy prediction for the assumption node Hyphotesis is 10.16%. Difference between the two prediction values is 20.34%.

Mean of the bankruptcy prediction for the sample nodes X_i in P-T discretizing method is 21.73%. Also in this method the bankruptcy prediction for the assumption node Hyphotesis is 5.84%. Difference between the two prediction values is 15.89%.

From these results we can conclude that the bankruptcy prediction has a higher degree of accuracy using the P-T discretizing method.

References

[1] Hogg, R.V., and A.T. Craig, *Introduction to Mathematical Statistics*, Macmillan, New York, 1972

[2] Cooper, G.F., *The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks*, Artificial Intelligence, Vol. 33, 1990

[3] Lindley, D.V., *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge University Press, London, 1985

[4] Berry, D.A., *Statistics: A Bayesian Perspective*, Wadsworth, Belmont, California, 1996

[5] Anghel Ion – *Falimentul – radiografie şi predicție*, Ed. Economică, Bucureşti, 2002

[6] Agena 2007, Press Release, *http://www.agenarisk.com/agenarisk/case_13.shtml*

[7] Crăciun Mihaela–Daciana, Bucerzan Dominic, Rațiu Crina, *Prediction Analysis of Bankrupcty using Bayesian Networks*, Analele Universității Maritime Constanța, Anul XI, Volumul 14, 2010

MIHAELA CRĂCIUN
"Aurel Vlaicu University" of Arad
Dep. of Mathematics and Computer Science
Str.Elena Dragoi, No.2, 310330 - Arad,
Complex Universitar M
ROMANIA
mihaeladacianacraciun@yahoo.com

DOMINIC BUCEZAN
"Aurel Vlaicu University" of Arad
Dep. of Mathematics and Computer Science
Str.Elena Dragoi, No.2, 310330 - Arad,
Complex Universitar M
ROMANIA
dominic@bbcomputer.ro

CRINA RAȚIU
DARAMEC SRL
Şofronea, Arad
ROMANIA
ratiu_anina@yahoo.com