Statistical Data Mining for Computational Financial Modeling

Ali Serhan KOYUNCUGIL, Ph.D.
Capital Markets Board of Turkey - Research Department
Ankara, Turkey
askoyuncugil@gmail.com
www.koyuncugil.org

Nermin OZGULBAS, Ph.D.
Baskent University - Department of Healthcare Management
Ankara, Turkey
ozgulbas@baskent.edu.tr

Overview of Financial Studies

- Financial ratios derived from firms' balance sheets and income statements have been using as most useful variables in financial studies.

- Financial ratios are used to
  - evaluate the overall financial condition,
  - measure financial performance,
  - identify risk and distress probability

- Analysts have been searching the **more efficient methodologies, statistical analysis, algorithms and models** to solve the problems of financial analysis especially by financial ratios.

Some Problems in Financial Analysis/Modeling

- Selecting statistically significant and financially meaningful ratios,
- Determining performance and risk indicators,
- Determining industrial (standard) ratios,
- Using operational and financial variables together,
- Detecting early warning signs for financial risks,
- Financial profiling and classification of the firms,
- Determining the financial road maps.

Objective

The objective of this study is presenting a computational financial model by data mining which is capable to solve the problems of financial analysis/modeling.

Financial Modelling - Discovery of Knowledge - Data Mining

- The identification of the factors for financial modelling by clarifying the relationship between the variables defines as the **discovery of knowledge**.

- Also, automated and prediction oriented information discovery process coincides the definition of **data mining**.

- Therefore, the ideal method for financial modeling is **data mining** that is started to be used more frequently nowadays for financial studies.

Data Mining

According to Koyuncugil&Ozgulbas, data mining is a collection of evolved statistical analysis, machine learning and pattern recognition methods via intelligent algorithms which are using for automated uncovering and extraction process of hidden predictional information, patterns, relations, similarities or dissimilarities in (huge) data.

Disciplines

Data mining is an intersection of

- Statistics,
- Machine learning,
- Pattern recognition,
- Databases,
- Artificial intelligence,
- Expert systems,
- Data Visulation,
- High speed computing,

etc. fields.

Data Mining Methods

In the scope of data mining methods;

- Linear and Logistic Regression,
- Discriminant Analysis,
- Cluster Analysis,
- Factor Analysis,
- Principal Component Analysis,
- Classification and Regression Trees (C&RT),
- CHi-Square Automatic Interaction Detector (CHAID),
- Association rules,
- K-nearest neighbour,
- (Artificial) Neural Networks,
- Self Organizing Maps (SOM),

can be count as principal methods.

Point of View

Data mining is an intersection of a lot disciplines but there are two integral parts of data mining as

- Information and Communication Technologies (ICT),
- Statistics.

Therefore, there are two main point of view of data mining as

- ICT
- Statistics

Statistical Data Mining

In statistical perspective,
Data Mining can be defined
as
Evolution of Statistical Analysis Methods
via
Intelligent Algorithms
For
Automated Prediction

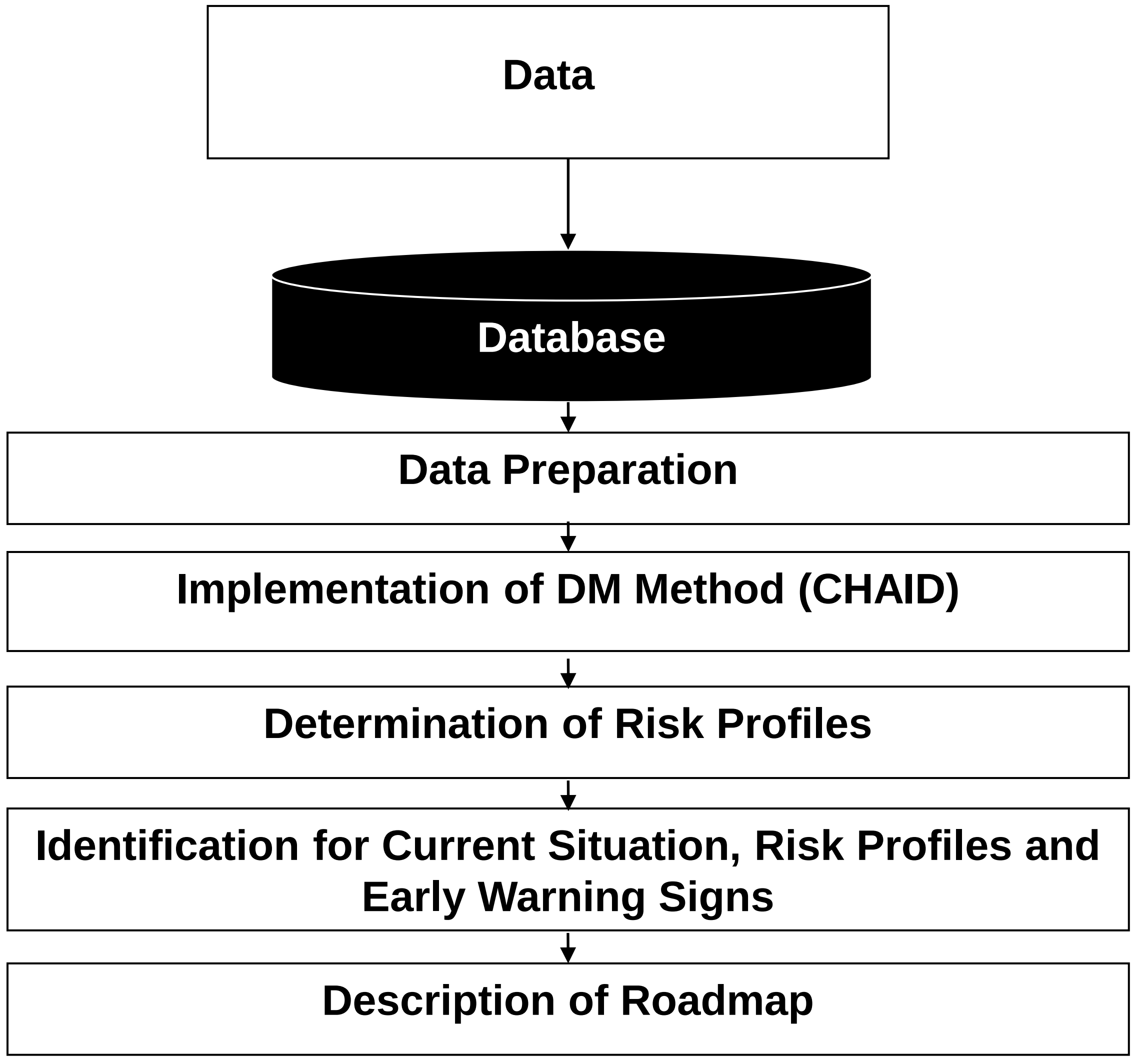
Goal of Data Mining

The only goal of Data Mining is
extracting
valuable high level knowledge
from
less informative data
(in context of huge data sets).

Data Mining for Financial Modelling

- This study is based on a Project which was funded by The Scientific and Technological Research Council of Turkey (TUBITAK).
- In this study, Chi-Square Automatic Interaction Detector (CHAID) decision tree algorithm has been used for financial modelling.
- Small and medium sized enterprises (SMEs) in Turkey were covered and their financial and operational data was used for mentioned purposes.
- This financial model could be use for
  – detecting financial and operational risk indicators,
  – determining financial risk profiles,
  – developing a financial early warning system (FEWS),
  – obtaining financial road maps for risk mitigation

Steps of the Model



I. Data Preparation

Data Sources:

- – Financial Data
- – Operational Data

Financial Data Preparation

- Financial data of SMEs was obtained from Turkish Central Bank (TCB) after permission.
- The study covered 7.853 SMEs’ data which was available from TCB in year 2007.
- Financial data that are gained from balance sheets and income statements was used to calculate financial indicators of system (Table 1).

Table 1. Some of Financial Ratios

Ratios	**Definition**
Return on Equity	Net Income / Total Assets
Return on Assets	Net Income/ Total Equity
Profit Margin	Net Income/ Total Margin
Equity Turnover Rate	Net Revenues / Equity
Total Assets Turnover Rate	Net Revenues / Total Assets
Inventories Turnover Rate	Net Revenues / Average Inventories
Fixed Assets Turnover Rate	Net Revenues / Fixed Assets
Tangible Assets to Long Term Liabilities	Tangible Assets / Long Term Liabilities
Days in Accounts Receivables	Net Accounts Receivable/ (Net Revenues /365)
Current Assets Turnover Rate	Net Revenues/ Current Assets
Fixed Assets to Long Term Liabilities	Fixed Assets / Long Term Liabilities
Tangible Assets to Equities	Tangible Assets /Equities
Long Term Liabilities to Constant Capital	Long Term Liabilities / Constant Capital
Long Term Liabilities to Total Liabilities	Long Term Liabilities / Total Liabilities
Current Liabilities to Total Liabilities	Current Liabilities / Total Liabilities
Total Debt to Equities	Total Debt / Equities
Equities to Total Assets	Total Equity/Total Assets
Debt Ratio	Total Dept/Total Assets
Current Account Receivables to Total Assets	Current Account Receivables/ Total Assets
Inventories to Current Assets	Total Inventories / Current Assets
Absolute Liquidity	(Cash+Banks+ Marketable Sec.+ Acc. Rec.) / Current Liab.
Quick Ratio (Liquidity Ratio)	(Cash+Marketable Sec.+ Acc. Rec.)/ Current Liab.
Current Ratio	Current Assets/ Current Liabilities

Operational Data Preparation

- Operational data (Table 2) which couldn’t be access by balance sheets and income statements for financial management requirements of SMEs collected via a field study in Ankara.

- A questionnaire designed for collecting data and data collected from Organized Industrial Region (OIR) of Ankara.

- The study covered 1,876 SMEs’ operational data in year 2007.

Table 2. Some of Operational Variables

- sector
- legal status
- number of partners
- number of employees
- annual turnover
- annual balance sheet
- financing model
- the usage situation of alternative financing
- technological infrastructure
- literacy situation of employees
- literacy situation of managers
- financial literacy situation of employees
- financial literacy situation of managers
- financial training need of employees
- financial training need of managers
- knowledge and ability levels of workers on financial administration
- financial problem domains
- current financial risk position of SMEs

Steps of Preparation of Data

- Calculation of financial indicators and collecting of operational indicators
- Reduction of repeating variables in different indicators to solve the problem of Collinearity / Multicollinearity
- Imputation of missing data
- Solution of outlier and extreme value problem

II. Implementation of Data Mining Method (CHAID)

A data mining method, **Chi-Square Automatic Interaction Detector (CHAID)** decision tree algorithm, was used in the study for modeling, financial profiling and developing FEWS.

CHAID

- CHAID algorithm organizes Chi-square independency test among the target variable and predictor variables, starts from branching the variable which has the strongest relationship and arranges statistically significant variables on the branches of the tree due to the strength of the relationship.

- CHAID has multi-branches, while other decision trees are branched in binary. Thus, all of the important relationships in data can be investigated until the subtle details.

III. Determination of Risk Profiles

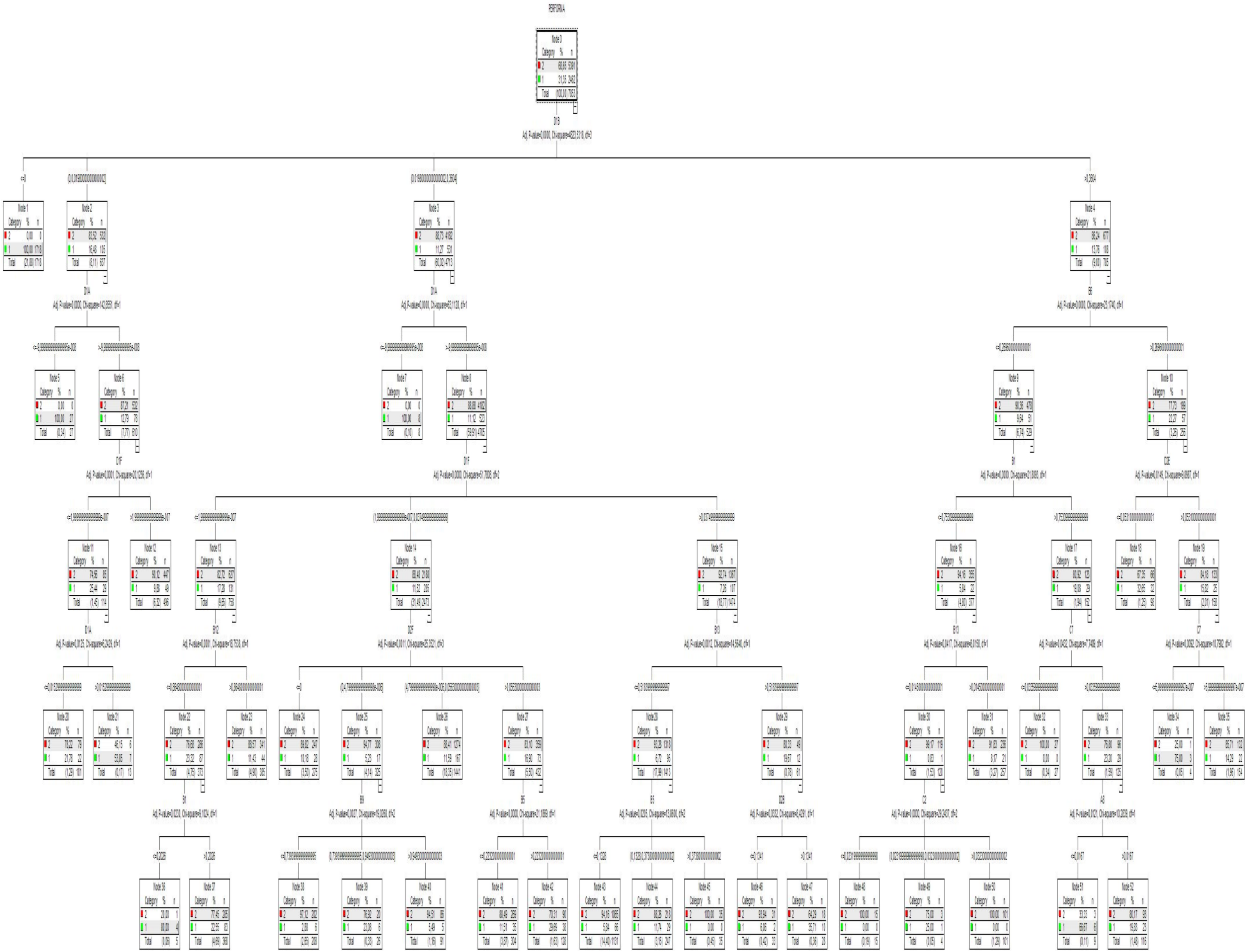
- In essence, the study identifies all the different risk profiles.
- Here the term risk means the risk that is caused because of the financial failures of enterprises.

Risk Profiles According to Financial Variables

- It was determined that 5.391 SMEs (68,65 %) had good financial performance, and 2.462 SMEs (31,38 %) had poor financial performance.

- SMEs were categorized into 31 different financial risk profiles

- 14 variables affected financial risk of SMEs.

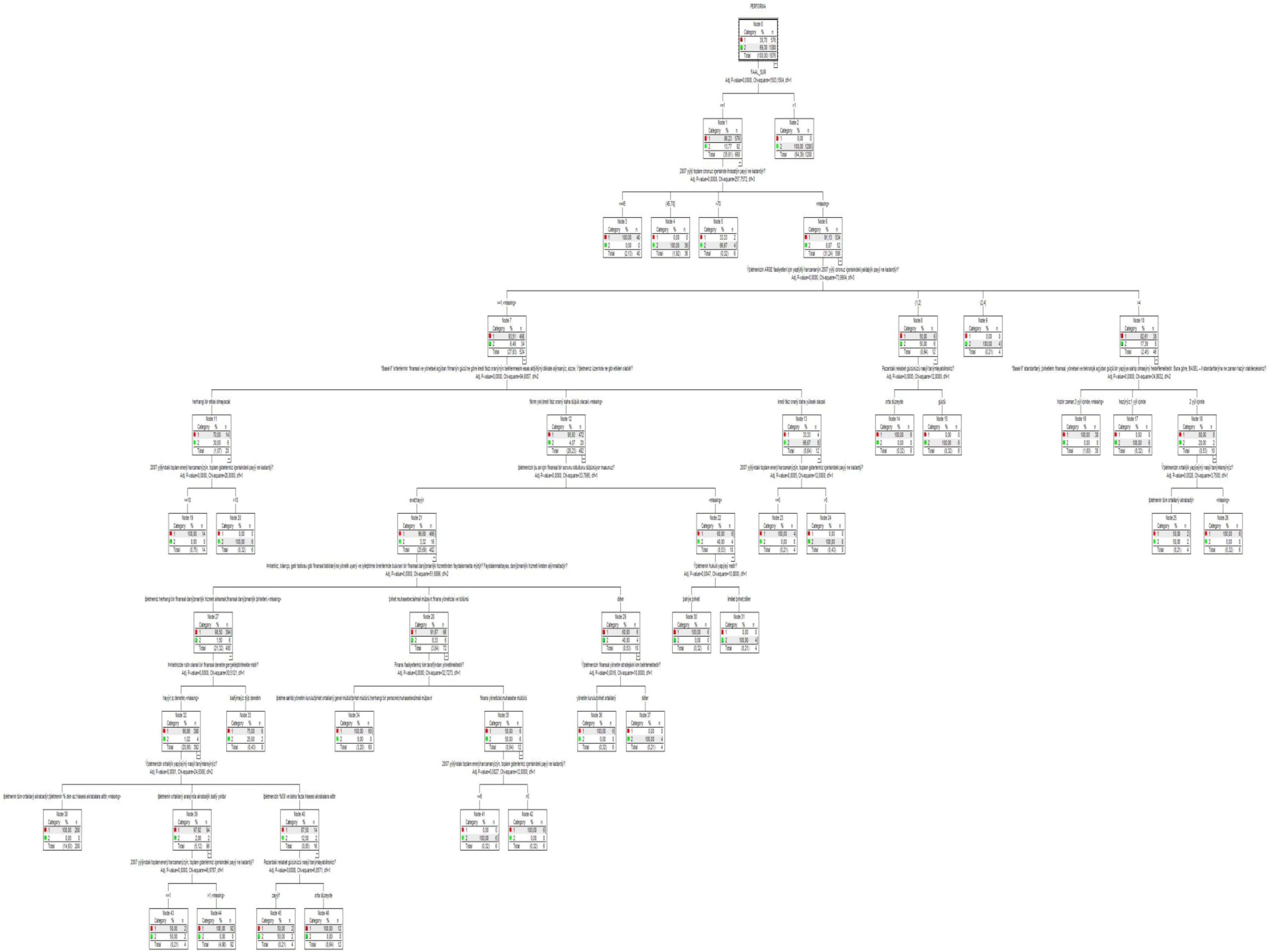

PERFORMA
Node 0
Category % n
2 68,65 5391
1 31,35 2462
Total (100,00) 7853
D1B
Adj. P-value=0,0000, Chi-square=4823,5318, df=3

Code	**Financial Indicators**
D1B	Profit Before Tax to Own Funds
D1A	Return on Equity (Net Profit to Own Funds)
D1F	Cumulative Profitability Ratio
B1	Total Loans to Total Assets
D2E	Operating Expenses to Net Sales
B12	Short-Term Liabilities to Total Loans
D2F	Interest Expenses to Net Sales
B13	Bank Loans to Total Assets
C7	Own Funds Turnover
B9	Fixed Assets to Long Term Loans+ Own Funds
B5	Long-Term Liabilities to Total Liabilities
D2B	Gross Profit to Net Sales
C2	Receivables Turnover
A8	Short-Term Receivables to Total Assets Total Assets
B6	Inventory Dependency Ratio

	Financial Indicators													
Nodes	**D1B**	**D1A**	**D1F**	**D2F**	**B12**	**B1**	**B9**	**B5**	**D2B**	**B6**	**B13**	**C7**	**A8**	**D2E**
0,1	≤ 0													
0,2,5	0-0,198	≤ 0												
0,2,6,11,20	0-0,198	0-0,015	≤0,0000002											
0,2,6,21	0-0,198	> 0,015	≤0,0000002											
0,2,6,12	0-0,198	> 0	> 0,0000002											
0,3,7	0,198-0,36	≤ 0												
0,3,8,13,22,36	0,198-0,36	> 0	≤0,0000002		≤ 0,86	≤ 0,20								
0,3,8,13,22,37	0,198-0,36	> 0	≤0,0000002		≤ 0,86	> 0,20								
0,3,8,13,23	0,198-0,36	> 0	≤0,0000002		>0,86									
0,3,8,14,24	0,198-0,36	> 0	0,0000002-0,04	≤ 0										
0,3,8,14,25,38	0,198-0,36	> 0	0,0000002-0,04	0-0,0000048			≤ 0,74							
0,3,8,14,25,39	0,198-0,36	> 0	0,0000002-0,04	0-0,0000048			0,74-0,95							
0,3,8,14,25,40	0,198-0,36	> 0	0,0000002-0,04	0-0,0000048			>0,95							
0,3,8,14,26	0,198-0,36	> 0	0,0000002-0,04	0,0000048-0,06										
0,3,8,14,27,41	0,198-0,36	> 0	0,0000002-0,04	>0,06				≤ 0,22						
0,3,8,14,27,42	0,198-0,36	> 0	0,0000002-0,04	>0,06				>0,22						
0,3,8,15,28,43	0,198-0,36	> 0	>0,04					≤ 0,14			≤ 0,52			
0,3,8,15,28,44	0,198-0,36	> 0	>0,04					0,14-0,38			≤ 0,52			
0,3,8,15,28,45	0,198-0,36	> 0	>0,04					>0,38			≤ 0,52			
0,3,8,15,29,46	0,198-0,36	> 0	>0,04						≤ 0,13		>0,52			
0,3,8,15,29,47	0,198-0,36	> 0	>0,04						>0,13		>0,52			
0,4,9,16,30,48	>0,36					≤ 0,75				≤ 0,26	≤ 0,015			
0,4,9,16,30,49	>0,36					≤ 0,75				≤ 0,26	≤ 0,015			
0,4,9,16,30,50	>0,36					≤ 0,75				≤ 0,26	≤ 0,015			
0,4,9,16,31	>0,36					≤ 0,75				≤ 0,26	>0,015			
0,4,9,17,32	>0,36					>0,75				≤ 0,26		≤ 0,03		
0,4,9,17,33,51	>0,36					>0,75				≤ 0,26		>0,03	≤ 0,02	
0,4,9,17,33,52	>0,36					>0,75				≤ 0,26			>0,02	
0,4,10,18	>0,36									>0,26				≤ 0,05
0,4,10,19,34	>0,36									>0,26		≤ 0,0000006		>0,05
0,4,10,19,35	>0,36									>0,26		> 0,0000006		

Risk Profiles According to Operational Variables

- It was determined that 1.300 SMEs (69,30 %) had good financial performance, and 576 SMEs (30,70 %) had poor financial performance.
- SMEs were categorized into 28 different financial risk profiles
- 14 operational variables affected financial risk of SMEs.


PERFORMA

Operational Variables	**p**
Activity Duration	<0,0001
Proportion of Export to Sales	<0,0001
Proportion of R&D Expenses to Sales	<0,0001
Ready to Basel- II	<0,0001
Power of Competition in Market	=0,0005
Knowledge About Basel-II	=0,053
Partnership Status	=0,0001
Proportion of Energy Expenses to Total Expenses	<0,0001
Awareness About Finance	<0,0001
Using Financial Consultant	<0,0001
Auditing	<0,0001
Person Responsible From Financial Management	<0,0001
Person Responsible from Financial Strategies	=0,0016
Legal Status	=0,0047

IV. Identification for Current Situation of SME from Risk profiles and Early Warning Signs

- 31.38 % of the covered SMEs financially distress.

Financial Signs

- There were 8 variables related with risk. These are:
  - Profit Before Tax to Own Funds
  - Return on Equity
  - Cumulative Profitability Ratio
  - Total Loans to Total Assets
  - Long-Term Liabilities to Total Liabilities
  - Inventory Dependency Ratio
  - Bank Loans to Total Assets
  - Own Funds Turnover

- According to profiles, risk profiles of SMEs were determined.
- Best Profiles that contained SMEs without risk were 19,22,26,29.
- Every firm tries to be in these Profiles.

V. Description of Roadmaps for SMEs (financial variables)

Rod Maps	Profiles	Probility of no risk	Financial Indicators							
			D1B Profit Before Tax to Own Funds	**D1A** Return on Equity	**D1F** Cumulative Profitability Ratio	**B1** Total Loans to Total Assets	**B5** Long-Term Liabilities to Total Liabilities	**B6** Inventory Dependency Ratio	**B13** Bank Loans to Total Assets	**C7** Own Funds Turnover
1	19	% 100	0,198-0,36	> 0	>0,04		>0,38		≤ 0,52	
2	22	% 100	>0,36			≤ 0,75		≤ 0,26	≤ 0,015	
3	24	% 100	>0,36			≤ 0,75		≤ 0,26	≤ 0,015	
4	26	% 100	>0,36			>0,75		≤ 0,26		≤ 0,03

Conclusions

Contributions of Model:

- Determination of financial position and performance
- Selection of statistically useful and financially meaningful ratios for performance measurement
- Detection of industrial (standard) ratios
- Determination of risk levels
- Detection of financial and operational risk factors
- Detection of early warning signs
- Using all kinds of variables together
- Roadmaps for risk reduction

Also SMEs Could Use This Model for:

- Prevention for financial distress
- Decrease the possibility of bankruptcy
- Decrease risk rate
- Efficient usage of financial resources
- By efficiency;
  - Increase the competition capacity
  - New potential for export,
  - Decrease the unemployment rate
  - More taxes for government
- Adaptation to BASEL II Capital Accord

Acknowledgment

- This study is based on a project which was funded by The Scientific and Technological Research Council of Turkey (TUBITAK).

Thank you very much …

You can download the presentation from

www.koyuncugil.org

Statistical Data Mining for Computational Financial Modeling

Ali Serhan KOYUNCUGIL, Ph.D.
Capital Markets Board of Turkey - Research Department
Ankara, Turkey
askoyuncugil@gmail.com
www.koyuncugil.org

Nermin OZGULBAS, Ph.D.
Baskent University - Department of Healthcare Management
Ankara, Turkey
ozgulbas@baskent.edu.tr