شبكة المعلومات الجامعية

بسم الله الرحمن الرحيم

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

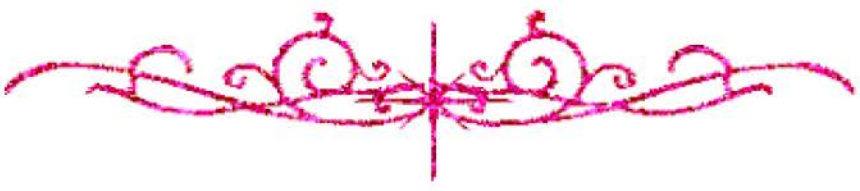
نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار

شبكة المعلومات الجامعية

بعض الوثائق الأصلية تالفة



بالرسالة صفحات لم ترد بالأصل

Cairo University
Institute of Statistical Studies and Research
Department of Computer and Information Science

A Tool For Knowledge Discovery In Life Assurance Data

By

Suzanne Soliman Ali Shafik

Thesis submitted in partial fulfillment of the requirements for Ph.D. in Computer and Information Science

Supervised by

Prof. Dr Ahmed Rafea
Computer science Dept.,
Faculty of Computers and Information
Cairo University

Dr Khaled Shaalan
Associate prof.
Computer science Dept.,
Faculty of Computers and Information
Cairo University

2003

Cairo University
Institute of Statistical Studies and Research
Computer Science and Information department.

Approval sheet

A tool for knowledge discovery In life assurance data

A thesis submitted in partial fulfillment of the
Ph.D. in Computer science
By

Suzanne Soliman Aly Shafeik

Computer Science and Information dept.
Institute of Statistical Studies and Research
Cairo University

2003

This thesis for PH.D. degree in computer sciences, department of computer science and information of statistical studies and research, Cairo University, has been approved by:

Name	Signature
Prof. Dr. Ahmed Rafea	.....
Prof. Dr. Fawzi Amer	.....
Prof. Dr. Mervat Gheith	.....
Dr. Khaled Shaalan	.....

Date : 20/7/2003

Acknowledgment

I would like to express my deep thanks and sincere appreciation to Prof. Dr. Ahmed Rafea to his valuable effort, continues support and expert guidance through all phases of this work.

My gratefully thanks, appreciation, and gratitude to Dr. khaled Shalaan for his fruitable effort, encouragement, support, and cooperation during all phases of the work.

My deep thanks to Prof. Dr. Mawaad Hasaneen Chairman of Misr Insurance Co. for his approval of using company life assurance historical data during the implementation phase.

My sincere thanks to IT department in Egyptian Insurance Supervisory Authority for providing me with life assurance market statistics in electronic form.

All my appreciation, thanks and gratitude to ISSR Computer and information science department for all kind provided facilities and cooperation.

My deep thanks, appreciation and gratitude to Prof. Dr. Adeeb Riyad Ghonemy to his support through valuable discussions, and who made all his library available during the research time.

Finally my deep sincere thanks to all my colleagues in IT sector in Misr Insurance Co. for their efforts and cooperation during implementation phase of this work.

Statement

I certify that this work has not been accepted in substance for any academic degree, and is not being concurrently submitted in candidature for any other degree. Any portions of this thesis for which I am indebted to other sources are mentioned and explicit references are given.

Student name : Suzanne Shafik

Table of Contents

Figures

Tables

Abstract

Many problems face the insurance companies due to the particularities of insurance industry. These problems are related to environmental factors, internal factors, and technical factors. Discovering relevant knowledge from available large volume of historical data and experts domain knowledge, could support the decision makers in achieving their goals.

This work uses prior domain knowledge to guide the mining of association rules in life assurance business environment. This approach is used in order to overcome the drawbacks of data mining using rule induction such as loss of information, discover too many obvious patterns, and mining of overwhelmed association rules. Life assurance data consist of enforced life assurance policies records, and many relevant historical transactional records. Each historical transaction has a transaction type associated with time stamp of occurrence, which represents an event to the assurance policy. These events compose an episode of events that have a certain pattern such as (emission, annulation, re-emission), (emission, amendment, loans, liquidation), or....etc. The episode of events is a subset of the transactions item-list. The final status of enforced policies in the company portfolio is affected dynamically by the sequential order of episode of events occurred during the policy life since its emission. This effect represents a positive or negative status on the company portfolio growth. It is required to handle dynamic nature of data, where appending a new event to an existing episode of events may change the status of life assurance policies, change discovered patterns and affect the support of mined rules. Two types of data are identified and used in mining rules, global data its attributes are time series of economic and demographic indicators, and insurance data aggregated on different levels and dimensions.

In order to help the decision maker a tool was built to discover knowledge at the macro and micro levels. The discovered rules, at the macro level, describe the impact of changes of national economic and demographic indicators on the growth of life assurance business. The discovered rules, at the micro level, describe the impact of different policies attributes, customer profiles, and market channels on company portfolio growth.

Our approach for data mining is a hybrid approach that consists of two data mining techniques namely: clustering and rule-induction. The clustering technique is applied on preprocessed multidimensional quantitative and categorical operational database