In 1st Balkan Conference on Informatics, 2003

Association Rules and Predictive Models for e-Banking Services

Vasilis Aggelis, Dimitris Christodoulakis

University of Patras
Department of Computer Engineering and Informatics
Rio, Patras, Greece
Vasilis.Aggelis@egnatiabank.gr
dxri@cti.gr

Abstract. The introduction of data mining methods in the banking area although conducted in a slower way than in other fields, mainly due to the nature and sensitivity of bank data, can already be considered of great assistance to banks as to prediction, forecasting and decision making. One particular method is the investigation for association rules between products and services a bank offers. Results are generally impressive since in many cases strong relations are established, which are not easily observed at a first glance. These rules are used as additional tools aiming at the continuous improvement of bank services and products helping also in approaching new customers. In addition, the development and continuous training of prediction models is a very significant task, especially for bank organizations. The establishment of such models with the capacity of accurate prediction of future facts enhances the decision making and the fulfillment of the bank goals, especially in case these models are applied on specific bank units. E-banking can be considered such a unit receiving influence from a number of different sides. Scope of this paper is the demonstration of the application of data mining methods to e-banking. In other words association rules concerning e-banking are discovered using different techniques and a prediction model is established, depending on e-banking parameters like the transactions volume conducted through this alternative channel in relation with other crucial parameters like the number of active users.

Keywords: Data Mining, e-Banking, Association Rules, Apriori, Generalized Rule Induction, Predictive models, Decision Trees, Evaluation charts

1. Introduction

Determination of association rules concerning bank data is a challenging though demanding task since:

- The volume of bank data is enormous. Therefore the data should be adequately prepared by the data miner before the final step of the application of the method.
- The objective must be clearly set from the beginning. In many cases not having a clear aim results in erroneous or no results at all.

- Good knowledge of the data is a prerequisite not only for the data miner but also for the final analyst (manager). Otherwise wrong results will be produced by the data miner and unreliable conclusions will be drawn by the manager.
- Not all rules are of interest. The analyst should recognize powerful rules as to decision making.

The challenge of the whole process rises from the fact that the relations established are not easily observed without use of data mining methods.

A number of data mining methods are nowadays widely used in treatment of financial and bank data [22]. The contribution of these methods in prediction making is essential. Study and prediction of bank parameter's behavior generally receives various influences. Some of these can be attributed to internal parameters while others to external financial ones. Therefore prediction making requires the correct parameter determination that will be used for predictions and finally for decision making. Methods like Classification, Regression, Neural Networks and Decision Trees are applied in order to discover and test predictive models.

Prediction making can be applied in specific bank data sets. An interesting data sector is e-banking. E-banking is a relatively new alternative payment and information channel offered by the banks. The familiarity of continuously more people to the internet use, along with the establishment of a feeling of increasing safety in conduction of electronic transactions imply that slowly but constantly e-banking users increase. Thus, the study of this domain is of great significance to a bank.

The increase of electronic transactions during the last years is quite rapid. E-banking nowadays offers a complete sum of products and services facilitating not only the individual customers but also corporate customers to conduct their transactions easily and securely. The ability to discover rules between different electronic services is therefore of great significance to a bank.

Identification of such rules offers advantages as the following:

- Description and establishment of the relationships between different types of electronic transactions.
- The electronic services become more easily familiar to the public since specific groups of customers are approached, that use specific payment manners.
- Customer approach is well designed with higher possibility of successful engagement.
- The improvement of already offered bank services is classified as to those used more frequently.
- Reconsidering of the usefulness of products exhibiting little or no contribution to the rules.

Of certain interest is the certification of the obtained rules using other methods of data mining to assure their accuracy.

This study is concerned with the identification of rules between several different ways of payment and also we focus in prediction making and testing concerning the transactions volume through e-banking in correlation to the number of active users. The software used is SPSS Clementine 7.0. Section 2 contains association rules' basic features and algorithms. A general description of prediction methods can be found in section 3, while in section 4 the process of investigation for association rules and the process of prediction making and testing is described. Experimental results can be found in section 5 and finally section 6 contains the main conclusions of this work accompanied with future work suggestions in this area.

2. Association rules basics

A rule consists of a left-hand side proposition (antecedent) and a right-hand side (consequent) [3]. Both sides consist of Boolean statements. The rule states that if the left-hand side is true, then the right-hand is also true. A probabilistic rule modifies this definition so that the right-hand side is true with probability p, given that the left-hand side is true.

A formal definition of association rule [6, 7, 14, 15, 18, 19, 21] is given below.

Definition. An association rule is a rule in the form of

$$X \rightarrow Y \tag{1}$$

Where *X* and *Y* are predicates or set of items.

As the number of produced associations might be huge, and not all the discovered associations are meaningful, two probability measures, called *support* and *confidence*, are introduced to discard the less frequent associations in the database. The support is the joint probability to find X and Y in the same group; the confidence is the conditional probability to find in a group Y having found X.

Formal definitions of support and confidence [1, 3, 14, 15, 19, 21] are given below.

Definition Given an itemset pattern X, its *frequency fr(X)* is the number of cases in the data that satisfy X. *Support* is the frequency fr(X $\wedge$ Y). *Confidence* is the fraction of rows that satisfy Y among those rows that satisfy X,

$$c(X \rightarrow Y) = \frac{fr(X \wedge Y)}{fr(X)} \tag{2}$$

In terms of conditional probability notation, the empirical accuracy of an association rule can be viewed as a maximum likelihood (frequency-based) estimate of the conditional probability that Y is true, given that X is true.

2.1 Apriori algorithm

Association rules are among the most popular representations for local patterns in data mining. Apriori algorithm [1, 2, 3, 6, 10, 14] is one of the earliest for finding association rules. This algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. This algorithm contains a number of passes over the database. During pass k, the algorithm finds the set of frequent itemsets L_k of length k that satisfy the minimum support requirement. The algorithm terminates when L_k is empty. A pruning step eliminates any candidate, which has a smaller subset. The pseudo code for Apriori Algorithm is following:

```
Cₖ: candidate itemset of size k
```

```
Lk: frequent itemset of size k

L1 = {frequent items};
For (k=1; Lk != null; k++) do begin
          Ck+1 = candidates generated from Lk;
          For each transaction t in database do
            Increment the count of all candidates in
            Ck+1 that are contained in t
          Lk+1 = candidates in Ck+1with min_support
          End
Return Lk;
```

2.2 Generalized Rule Induction (GRI)

In the generalized rule induction task [3, 4, 5, 11, 15, 16, 17], the goal is to discover rules that predict the value of a goal attribute, given the values of other attributes. However the goal attribute can be any attribute not occurring in the antecedent of the rule.

3. Classification, and Regression

There are two types of problems that can be solved by Classification and Regression methods [3].

Regression-type problems. The problem of regression consists of obtaining a functional model that relates the value of a response continuous variable Y with the values of variables $X_1, X_2, ..., X_v$ (the predictors). This model is usually obtained using samples of the unknown regression function.

Classification-type problems. Classification-type problems are generally those where one attempts to predict values of a categorical response variable from one or more continuous and/or categorical predictor variables.

3.1 Classification and Regression Trees (C&RT)

C&RT method [23, 24] builds classification and regression trees for predicting continuous dependent variables (regression) and categorical predictor variables (classification). There are numerous algorithms for predicting continuous variables or categorical variables from a set of continuous predictors and/or categorical factor effects. In most general terms, the purpose of the analysis via tree-building algorithms is to determine a set of *if-then* logical conditions that permit accurate prediction or classification of cases.

Tree classification techniques [3, 25, 26], when they "work" and produce accurate predictions or predicted classifications based on a few logical if-then conditions, have a number of advantages over many of those alternative techniques.

Simplicity of results. In most cases, the interpretation of results summarized in a tree is very simple. This simplicity is useful not only for purposes of rapid classification of new observations but can also often yield a much simpler "model" for explaining why observations are classified or predicted in a particular manner.

Tree methods are nonparametric and nonlinear. The final results of using tree methods for classification or regression can be summarized in a series of logical if-then conditions (tree nodes). Therefore, there is no implicit assumption that the underlying relationships between the predictor variables and the response variable are linear, follow some specific non-linear link function, or that they are even monotonic in nature. Thus, tree methods are particularly well suited for data mining tasks, where neither a priori knowledge is available nor any coherent set of theories or predictions regarding which variables are related and how. In those types of data analysis, tree methods can often reveal simple relationships between just a few variables that could have easily gone unnoticed using other analytic techniques.

3.2 Evaluation Charts

Comparison and evaluation of predictive models in order for the best to be chosen is a task easily performed by the evaluation charts [4, 27]. They reveal the performance of the models concerning particular outcomes. They are based on sorting the records as to the predicted value and confidence of the prediction, splitting the records into groups of equal size (quantiles), and finally plotting the value of the business criterion for each quantile, from maximum to minimum.

Different types of evaluation charts emphasize on a different evaluation criteria.

Gains Charts: The proportion of total hits that occur in each quantile are defined as Gains. They are calculated as the result of: (number of hits in quantile/total number of hits) X 100%.

Lift Charts:Lift is used to compare the percentage of records occupied by hits with the overall percentage of hits in the training data. Its values are computed as (hits in quantile/records in quantile) / (total hits/total records).

Evaluation charts can also be expressed in cumulative form, meaning that each point equals the value for the corresponding quantile plus all higher quantiles. Cumulative charts usually express the overall performance of models in a more adequate way, whereas non-cumulative charts [4] often succeed in indicating particular problem areas for models.

The interpretation of an evaluation chart is a task certainly depended on the type of chart, although there are some characteristics common to all evaluation charts. Concerning cumulative charts, higher lines indicate more effective models, especially on the left side of the chart. In many cases, when comparing multiple models lines cross, so that one model stands higher in one part of the chart while another is elevated higher than the first in a different part of the chart. In this case, it is necessary to consider which portion of the sample is desirable (which defines a point on the x axis) when deciding which model is the appropriate.

Most of the non-cumulative charts will be very similar. For good models, noncumulative charts [4] should be high toward the left side of the chart and low toward the right side of the chart. Dips on the left side of the chart or spikes on the right side can indicate areas where the model is predicting poorly. A flat line across the whole graph indicates a model that essentially provides no information.

Gains charts. Cumulative gains charts extend from 0% starting from the left to 100% at the right side. In the case of a good model, the gains chart rises steeply towards 100% and then levels off. A model that provides no information typically follows the diagonal from lower left to upper right.

Lift charts. Cumulative lift charts tend to start above 1.0 and gradually descend until they reach 1.0 heading from left to right. The right edge of the chart represents the entire data set, so the ratio of hits in cumulative quantiles to hits in data is 1.0. In case of a good model the lift chart, lift should start well above 1.0 on the left, remain on a high plateau moving to the right, trailing off sharply towards 1.0 on the right side of the chart. For a model that provides no information, the line would hover around 1.0 for the entire graph.

4. Investigation for association rules and Generation of predictions in e-banking data set

In this section the determination of relations between different payment types e-banking users conduct is described. These relations are of great importance to the bank as they enhance the selling of its products, by making the bank more aware of its customer's behavior and the payment way they prefer. They also contribute in improvement of its services.

Namely the following types of payments were used:

1. *DIASTRANSFER Payment Orders* – Funds Transfers between banks that are participating in DIASTRANSFER System.
2. *SWIFT Payment Orders* – Funds Transfers among national and foreign banks.
3. *Funds Transfers* – Funds Transfers inside the bank.
4. *Forward Funds Transfers* – Funds Transfers inside the bank with forward value date.
5. *Greek Telecommunications Organization (GTO) and Public Power Corporation (PPC) Standing Orders* – Standing Orders for paying GTO and PPC bills.
6. *Social Insurance Institute (SII) Payment Orders* – Payment Orders for paying employers' contributions.
7. *VAT Payment Orders.*
8. *Credit Card Payment Orders.*

For the specific case of this study a random sample of e-banking customers, both individuals and companies, was used. A Boolean value is assigned to each different payment type depending on whether the payment has been conducted by the users or not. In case the user is individual field «company» receives the value 0, else 1.

A sample of the above data set is shown in Table 1.

User	**Comp**	**Dias Transfer P.O.**	**Swift P.O.**	**Funds Transfer**	**Forward Funds Transfer**	**GTO-PPC S.O.**	**SII P.O.**	**VAT P.O.**	**Cr. Card P.O.**
…	…	…	…	…	…	…	…	…	…

User103	0	T	F	T	T	F	F	F	T
User104	1	T	T	T	T	F	T	T	F
User105	1	T	F	T	T	T	T	T	T
User106	0	T	F	T	F	T	F	F	T
…	…	…	…	…	…	…	…	…	…

Table 1.

In total the sample contains 1959 users.

In order to discover association rules the Apriori [4] and Generalized Rule Induction (GRI) [4] methods were used. These rules are statements in the form *if antecedent then consequent*.

In the case of this study, as stated above, another objective is the production and test of predictions about the volume of e-banking transactions in relation to the active users. The term financial transactions stands for all payment orders or standing orders a user carries out, excluding transactions concerning information content like account balance, detailed account transactions or mini statement. The term «active» describes the user who currently makes use of the electronic services a bank offers. Active users are a subgroup of the enlisted users.

One day is defined as the time unit. The number of active users is the predictor variable (Count_Of_Active_Users) while the volume of financial transactions is assumed to be the response variable (Count_Of_Payments).

A sample of the above data set is shown in Table 2

Transaction Day	Count_Of_Active_Users	Count_Of_Payments
...	...	...
27/8/2002	99	228
28/8/2002	107	385
29/8/2002	181	915
30/8/2002	215	859
...	...	...

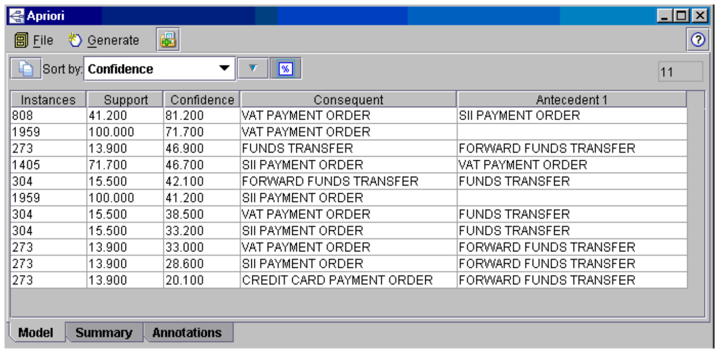
Table 2.

The date range for which the data set is applicable counts from April 20th, 2001 until December 12th, 2002. Data set includes data only for active days (holydays and weekends not included), which means 387 occurrences.

In order to create predictions Classification and Regression Tree (C&R Tree) [4] was used.

5. Experimental Results

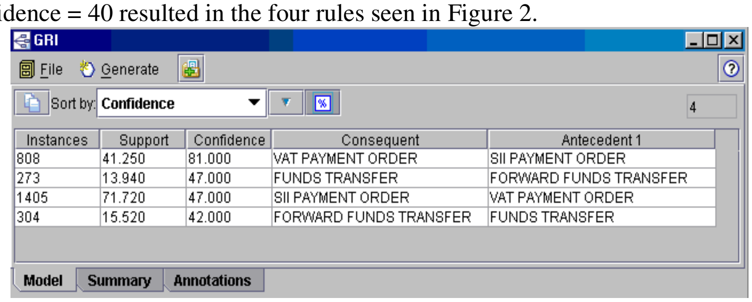
Apriori: By the use of Apriori method and setting Minimum Rule Support = 10 and Minimum Rule Confidence = 20 eleven rules were determined and shown in Figure 1.



Instances	Support	Confidence	Consequent	Antecedent 1
808	41.200	81.200	VAT PAYMENT ORDER	SII PAYMENT ORDER
1959	100.000	71.700	VAT PAYMENT ORDER	
273	13.900	46.900	FUNDS TRANSFER	FORWARD FUNDS TRANSFER
1405	71.700	46.700	SII PAYMENT ORDER	VAT PAYMENT ORDER
304	15.500	42.100	FORWARD FUNDS TRANSFER	FUNDS TRANSFER
1959	100.000	41.200	SII PAYMENT ORDER	
304	15.500	38.500	VAT PAYMENT ORDER	FUNDS TRANSFER
304	15.500	33.200	SII PAYMENT ORDER	FUNDS TRANSFER
273	13.900	33.000	VAT PAYMENT ORDER	FORWARD FUNDS TRANSFER
273	13.900	28.600	SII PAYMENT ORDER	FORWARD FUNDS TRANSFER
273	13.900	20.100	CREDIT CARD PAYMENT ORDER	FORWARD FUNDS TRANSFER

Fig. 1.

GRI: Using method GRI with Minimum Rule Support = 10 and Minimum Rule Confidence = 40 resulted in the four rules seen in Figure 2.



Instances	Support	Confidence	Consequent	Antecedent 1
808	41.250	81.000	VAT PAYMENT ORDER	SII PAYMENT ORDER
273	13.940	47.000	FUNDS TRANSFER	FORWARD FUNDS TRANSFER
1405	71.720	47.000	SII PAYMENT ORDER	VAT PAYMENT ORDER
304	15.520	42.000	FORWARD FUNDS TRANSFER	FUNDS TRANSFER

Fig. 2.

After comparison between the rules obtained by the two methods it can be concluded that the most powerful is:

– *If* SII Payment Order *then* VAT Payment Order (confidence=81). (Rule 1)

Additionally it can be observed that this rule is valid for the reverse case too, namely:

– *If* VAT Payment Order *then* SII Payment Order (confidence= 47). (Rule 2)

Two other rules obtained exhibiting confidence > 40 are the following:

– *If* Funds Transfers *then* Forward Funds Transfers (confidence=47), (Rule 3) valid also reversely.

– *If* Forward Funds Transfers *then* Funds Transfers (confidence=42), (Rule 4)

As seen, there exists strong relationship between VAT Payment Order and SII Payment Order, as is the case between Funds Transfers and Forward Funds Transfers.

These strong relationships are visualized on the web graph of Figure 3 and the tables of its links in Figure 4.

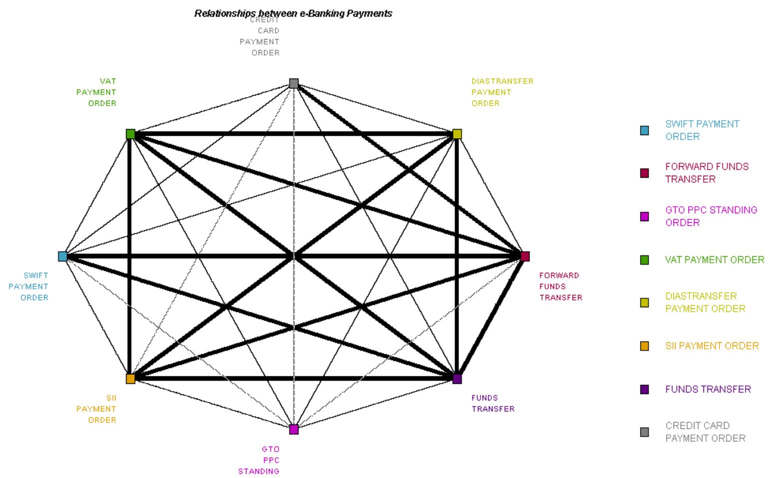


Fig. 3.

Strong Links

Links	Field 1	Field 2
656	VAT PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
128	FUNDS TRANSFER = "T"	FORWARD FUNDS TRANSFER = "T"
117	FUNDS TRANSFER = "T"	VAT PAYMENT ORDER = "T"
101	FUNDS TRANSFER = "T"	SII PAYMENT ORDER = "T"
90	FORWARD FUNDS TRANSFER = "T"	VAT PAYMENT ORDER = "T"
78	FORWARD FUNDS TRANSFER = "T"	SII PAYMENT ORDER = "T"
55	FORWARD FUNDS TRANSFER = "T"	CREDIT CARD PAYMENT ORDER = "T"
47	SWIFT PAYMENT ORDER = "T"	FORWARD FUNDS TRANSFER = "T"
40	DIASTRANSFER PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
37	DIASTRANSFER PAYMENT ORDER = "T"	FUNDS TRANSFER = "T"
37	SWIFT PAYMENT ORDER = "T"	FUNDS TRANSFER = "T"
36	DIASTRANSFER PAYMENT ORDER = "T"	VAT PAYMENT ORDER = "T"

Medium Links

Links	Field 1	Field 2
33	SWIFT PAYMENT ORDER = "T"	VAT PAYMENT ORDER = "T"
31	SWIFT PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
27	FUNDS TRANSFER = "T"	CREDIT CARD PAYMENT ORDER = "T"
26	CREDIT CARD PAYMENT ORDER = "T"	VAT PAYMENT ORDER = "T"
25	DIASTRANSFER PAYMENT ORDER = "T"	FORWARD FUNDS TRANSFER = "T"
25	DIASTRANSFER PAYMENT ORDER = "T"	SWIFT PAYMENT ORDER = "T"
24	VAT PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"
17	SII PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"
17	FUNDS TRANSFER = "T"	GTO PPC STANDING ORDER = "T"
17	SWIFT PAYMENT ORDER = "T"	CREDIT CARD PAYMENT ORDER = "T"
17	DIASTRANSFER PAYMENT ORDER = "T"	CREDIT CARD PAYMENT ORDER = "T"
15	DIASTRANSFER PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"

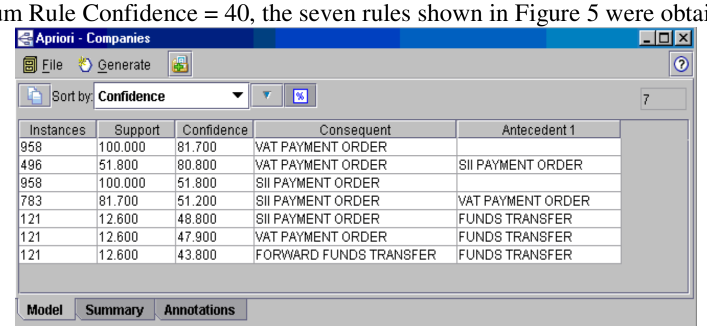
Weak Links

Links	Field 1	Field 2
13	CREDIT CARD PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
10	CREDIT CARD PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"
7	FORWARD FUNDS TRANSFER = "T"	GTO PPC STANDING ORDER = "T"
6	SWIFT PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"

Fig. 4.

Interestingness [8, 9, 12, 13, 19, 20] of the above rules is high bearing in mind that generally a pattern or rule can be characterized interesting if it is easily understood, unexpected, potentially useful and actionable or it validates some hypothesis that a user seeks to confirm. Given the fact that payment types used in the rules exhibiting the highest confidence (VAT Payment Order, SII Payment Order) are usually payments conducted by companies and freelancers, a smaller subdivision of the whole sample was considered containing only companies (958 users) for which the methods Apriori και GRI were applied. The results are described below.

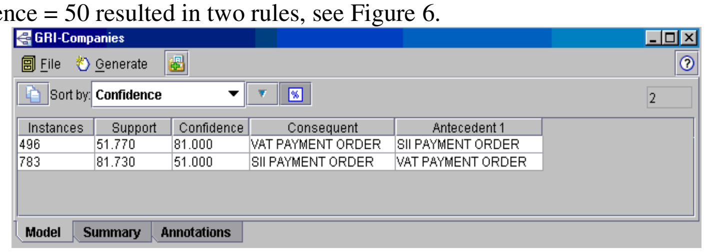
Apriori: Using Apriori Method and setting Minimum Rule Support = 10 and Minimum Rule Confidence = 40, the seven rules shown in Figure 5 were obtained.



Instances	Support	Confidence	Consequent	Antecedent 1
958	100.000	81.700	VAT PAYMENT ORDER	
496	51.800	80.800	VAT PAYMENT ORDER	SII PAYMENT ORDER
958	100.000	51.800	SII PAYMENT ORDER	
783	81.700	51.200	SII PAYMENT ORDER	VAT PAYMENT ORDER
121	12.600	48.800	SII PAYMENT ORDER	FUNDS TRANSFER
121	12.600	47.900	VAT PAYMENT ORDER	FUNDS TRANSFER
121	12.600	43.800	FORWARD FUNDS TRANSFER	FUNDS TRANSFER

Fig. 5.

GRI: GRI method with Minimum Rule Support = 10 and Minimum Rule Confidence = 50 resulted in two rules, see Figure 6.



Instances	Support	Confidence	Consequent	Antecedent 1
496	51.770	81.000	VAT PAYMENT ORDER	SII PAYMENT ORDER
783	81.730	51.000	SII PAYMENT ORDER	VAT PAYMENT ORDER

Fig. 6.

The comparison of the above rules leads to the conclusion that the validity of rules 1 and 2 is confirmed exhibiting even increased support:

– *If* SII Payment Order *then* VAT Payment Order (confidence=81). (Rule 5)
– *If* VAT Payment Order *then* SII Payment Order (confidence= 51). (Rule 6)

5.1 CR&T

In Figure 7 the conditions defining the partitioning of data discovered by the algorithm C&RT are displayed. These specific conditions compose a predictive model.

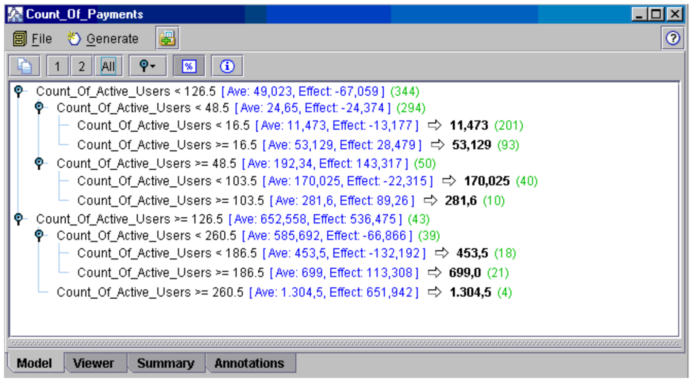


Fig. 7.

C&RT algorithm [4] works by recursively partitioning the data based on input field values. The data partitions are called *branches*. The initial branch (sometimes called the *root*) encompasses all data records. The root is split into subsets or *child branches*, based on the value of a particular input field. Each child branch may be further split into sub-branches, which may in turn be split again, and so on. At the lowest level of the tree are branches that have no more splits. Such branches are known as *terminal branches,* or *leaves.*

Figure 7 shows the input values that define each partition or branch and a summary of output field values for the records in that split.

For example a condition with many instances (201) is:

If Count_of_Active_Users < 16.5 then Count_of_Payments ⇨ 11,473,

meaning that if the number of active users during a working day is less than 16.5 then the number of transactions is predicted approximately at the value of 11.

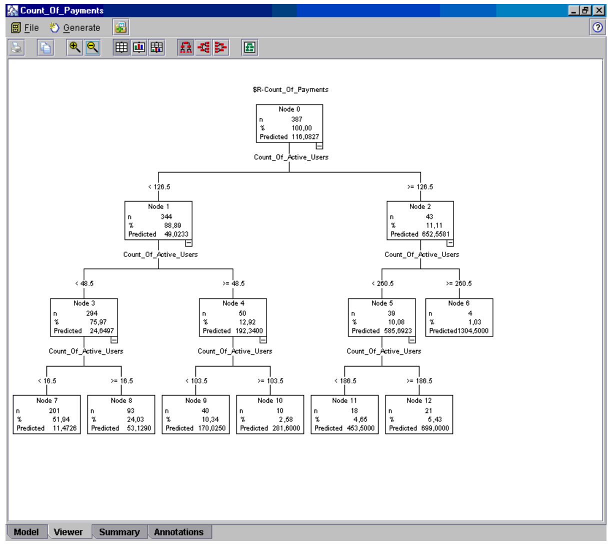


Fig. 8.

Figure 8 shows a graphical display of the structure of the tree in detail.

Goodness of fit of the discovered model was assessed by the use of evaluation charts. Examples of evaluation Charts (Figure 9 to Figure 10) are shown below.

Gains Chart: This Chart (Figure 9) shows that the gain rises steeply towards 100% and then levels off. Using the prediction of the model, the percentage of Predicted Count of Payments for the percentile is calculated and these points create the lift curve. An important notice is that the efficiency of a model is increased with the area between the lift curve and the baseline.

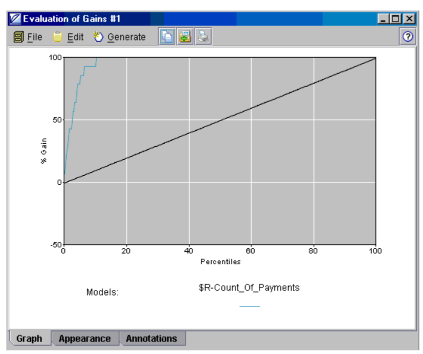


Fig. 9.

Lift Chart: As can be seen in Figure 10, Chart starts well above 1.0 on the left, remains on a high plateau moving to the right, and then trails off sharply towards 1.0 on the right side of the chart. Using the prediction of the model shows the actual lift.

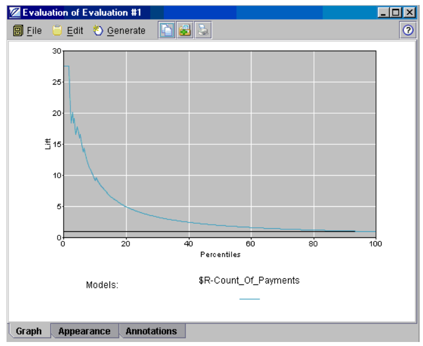


Fig. 10.

In Figure 11 the Count of Payment and the Predicted Count of Payment as functions of the number of active users can be seen. The Predicted Count of Payment line clearly reveals the partitioning of the data due to the application of the C&RT algorithm.

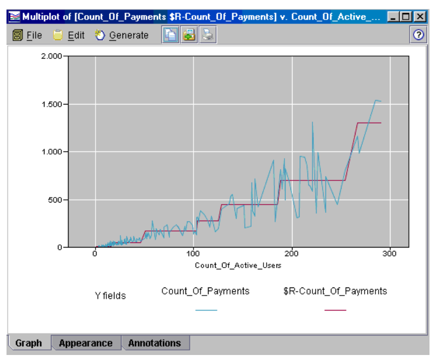


Fig. 11.

6. Conclusions and Future work

In the present paper the way of discovery of association rules between different types of e-banking payment offered by a bank is described along with experimental results.

The basic outcome is that VAT Payment Orders and SII Payment Orders are the most popular and interconnected strongly. The detection of such relationships offers a bank a detailed analysis that can be used as a reference point for the conservation of the customer volume initially but also for approaching new customer groups (companies, freelancers) who conduct these payments in order to increase its customer number and profit.

Similar patterns can be derived after analysis of payment types of specific customer groups resulting from various criteria either gender, residence area, age, profession or any other criteria the bank assumes significant. The rules to be discovered show clearly the tendencies created in the various customer groups helping the bank to approach these groups as well as re-design the offered electronic services in order to become more competitive.

In addition to the two methods used, detection of association rules can employ several other data mining methods such as decision trees in order to confirm the validity of the results.

In this study the establishment of a prediction model concerning the number of payments conducted through Internet as related to the number of active users is investigated along with testing its accuracy.

Using the methods C&RT it was concluded that there exists strong correlation between the number of active users and the number of payments these users conduct. It is clear that the increase of the users' number results in increase of the transactions made. Therefore, the enlargement of the group of active users should be a strategic target for a bank since it increases the transactions and decreases its service cost. It has been proposed by certain researches that the service cost of a transaction reduces from €1.17 to just €0.13 in case it is conducted in electronic form. Therefore, a goal of the e-banking sector of a bank should be the increase of the proportion of active users compared to the whole number of enlisted customers from which a large number do not use e-services and is considered inactive.

Future work includes the creation of prediction models concerning other e-banking parameters like the transactions value, the use of specific payment types, commissions of electronic transactions and e-banking income in relation to internal bank issues. Also extremely interesting is the case of predictive models based on external financial parameters.

The prediction model of this study could be determined using also other methods of data mining. Use of different methods offers the ability to compare between them and select the more suitable. The acceptance and continuous training of the model using the appropriate Data Mining method results in extraction of powerful conclusions and results.

7. References

1. M.J.Zaki, S. Parthasarathy, W. Li, and M. Ogihara. "*Evaluation of Sampling for Data Mining of Association Rules*", 7th Workshop Research Iss. Data Engg., 1997
2. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A.I. Verkamo. "*Fast discovery of associations rules*", Advances in Knowledge Discovery and Data Mining, 1996
3. D. Hand, H. Mannila, and P. Smyth "*Principles of Data Mining*". The MIT Press, 2001.
4. "*Clementine 7.0 Users's Guide*". Integral solutions Limited, 2002.
5. A. Freitas. "*A Genetic Programming Framework for Two Data Mining Tasks: Classification and Generalized Rule Induction*", Genetic Programming 1997: Proceedings of the Second Annual Conference, 1997

6. E. Han, G. Karypis, and V. Kumar. "*Scalable Parallel Data Mining for Association Rules*", Proceedings of ACM SIGMOD Intl. Conf. on Management of Data, 1997
7. S. Brin, R. Motwani, and C. Silverstein. "*Beyond Market Baskets: Generalizing Association Rules to Correlations*", Proceedings ACM SIGMOD Conf. on Management of Data, 1997
8. M. Chen, J. Han, and P. Yu. "*Data Mining: An Overview from Database Perspective*", Ieee Trans. On Knowledge And Data Engineering, 1997
9. N. Ikizler, and H.A. Guvenir. "*Mining Interesting Rules in Bank Loans Data*",: Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, 2001.
10. R. Ng, L. Lakshmanan, and J. Han. "*Exploratory Mining and Pruning Optimizations of Constrained Association Rules*", Proceedings ACM SIGMOD International Conference on Management of Data, 1998
11. A. Freitas. "*A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery*", A. Ghosh and S. Tsutsui. (Eds.) Advances in Evolutionary Computation Springer-Verlag, 2002
12. A. Freitas. "*On Rule Interestingness Measures*", Knowledge-Based Systems journal, 1999
13. R. Hilderman, and H. Hmailton. "*Knowledge Discovery and Interestingness Measures: A Survey*", Technical Report CS 99-04, Department of Computer Science,
University of Regina ,1999
14. J. Hipp, U. Guntzer, and G. Nakhaeizadeh. "*Data Mining of Associations Rules and the Process of Knowledge Discovery in Databases*", 2002
15. A. Michail. "*Data Mining Library Reuse Patterns using Generalized Association Rules*", Proceedings of the 22nd International Conference on on Software Engineering , 2000.
16. J. Liu, and J. Kwok. "*An Extended Genetic Rule Induction Algorithm*", Proceedings of the Congress on Evolutionary Computation (CEC), 2000.
17. P. Smyth, and R. Goodman. "*Rule Induction using Information Theory*", Knowledge Discovery in Databases 1991, 1991
18. T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama. "*Data Mining Using Two-Dimensional Optimized Association Rules: Scheme, Algorithms, and Visualization*", Proceedings of the ACM SIGMOD International Conference on Management of Data, 1996.
19. R. Srikant, and R. Agrawal. "*Mining Generalized Association Rules*", Proceedings of the 21st Int'l Conference on Very Large Databases, 1995
20. T. Yilmaz, and A. Guvenir. "*Analysis and Presentation of Interesting Rules*", Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, 2001.
21. H. Toivonen. "*Discovery of Frequent Patterns in Large Data Collections*", Technical Report A-1996-5, Department of Computer Science, University of Helsinki, 1996
22. D. Foster, and R. Stine. "*Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy*", Center for Financial Institutions Working Papers from Wharton School Center for Financial Institutions, University of Pennsylvania, 2002.
23. J. Galindo. "*Credit Risk Assessment using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications*", Computational Economics Journal, 1997
24. R.Lewis. "*An Introduction to Classification and Regression tree (CART) Analysis*", Annual Meeting of the Society for Academic Emergency Medicine in San, 2000.
25. L. Breiman, J. H. Friedman, R.A. Olshen, and C.J. Stone. "*Classification and Regression Trees*", Wadsworth Publications, 1984
26. A. Buja, and Y. Lee. "*Data Mining Criteria for Tree-based Regression and Classification*", The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001
27. S.J. Hong, and S. Weiss. "*Advances in Predictive Model Generation for Data Mining*" Proceedings 1st International Workshop Machine Learning and Data Mining in Pattern Recognition, 1999