Data Mining Solutions for Local Municipalities

Gözde Bakırlı[1], Derya Birant[2], Erol Mutlu[1], Alp Kut[2], Levent Denktaş[1] and Dilşah Çetin[1]
[1]OLGU Computer Systems, Izmir, Turkey
[2]Dokuz Eylul University, Department of Computer Engineering, Izmir, Turkey
gozde.bakirli@olgu.com.tr
derya@cs.deu.edu.tr
erol.mutlu@olgu.com.tr
alp@cs.deu.edu.tr
levent.denktas@olgu.com.tr
dilsah.cetin@olgu.com.tr

Abstract: This study proposes data mining solutions for local municipalities to make their decision support mechanism easier. The purpose of this study is to get intelligent solutions related to local government services from past data and to estimate the future activities. It covers socio-cultural analyses, income/expense analyses, infrastructure analyses, fraud detection analyses, simplification, verification and similarity analyses. Proposed system is based on service oriented architecture. The purposes of this project are; to give information about current state, to facilitate decision making for future activities, to increase income and decrease expense, to supply easy and correct data input to the system and to supply easier document tracking system. Seventeen scenarios were created initially. These scenarios are; Staff Analyzing, Classifying Citizens According to Real Estate Tax, Distribution of Citizens delaying Real Estate Tax, Income Operations Analyzing, Fuel Oil Analyzing, Electricity Consumption Analyzing, Cash Desk Analyzing, Distribution of Corporate Foundation, Moveable Material Analyzing, Logs Analyzing, Water Notice Analyzing, User Accounts Analyzing, Accountancy Analyzing, Employee Analyzing, Estimation of Wages, Citizen Analyzing and Corporate Foundation Analyzing. Service Oriented Architecture (SOA) is used as software architecture. Five services - Association Rule Mining Web Service (ARMWS), Outlier Detection Analysis Web Service (ODAWS), Classification Web Service (CWS), Clustering Web Service (ClustWS) and Data Preparation Web Service (DPWS) - were created. 7 scenarios used ARMWS, 3 scenarios used ODAWS, 2 scenarios used CWS and ClustWS is used by 5 scenarios.

Keywords: data mining, applications of local government, structure and urban informatics, service oriented architecture

1. Introduction

Data mining is the process of extracting hidden patterns from large datasets in human understandable structure. Data mining has ability to tell important things that you didn't know or what is going to happen next. The main data mining methods are classification, clustering, outlier detection analysis and association rule learning.

Current municipal management information systems support just execution of operational processes and creation of standard reports. New method and techniques that include data mining should be used to utilize retrospective knowledge, to develop solutions which are close to human intelligence and to make prediction for the future.

Data mining solutions for local municipalities provide that local municipalities can discover hidden patterns, relationships, changes, irregularities, and rules from large datasets. Local municipalities can perform their decision making process more rational, more accurate and faster. Local government personnel can discover knowledge without having any theoretic information about data mining, by using developed business intelligence software.

The need to business intelligence increases. The more increase in data in Local Municipalities, the more need to the data mining. Increase in the prevalence of use of municipal management information systems and formation of huge data heaps that ordinary analyses are inadequate for, increases business intelligence need at municipal sector.

It simplifies the way to give information to the local government about current status and to make decision for the future. The usage of new technologies at municipal sector will become widespread. Additionally, it contributes to the national economy with income/expense, wastage and abuse analyses at local municipalities.

Bakırlı, G., Birant, D, Mutlu, E., Kut, A., Denktaş, L. and Çetin, D."Reference this paper as: "Data Mining Solutions for Local Municipalities" *Electronic Journal of e-Government* Volume 10 Issue 2 2012, (pp97 - 106), available online at www.ejeg.com

This study proposes to get intelligent solutions related to local government services from past data and to estimate the future activities with OMIS-DM (OLGU Management Information System-Data Mining). The aim of this study is to facilitate decision making for future activities and to give information about current situation of the local municipalities.

2. Related works

In recent years, some studies that data mining has been applied for municipalities have been done. Poles and Margonari (2009) developed a data mining tool which is called modeFrontier for Italian municipalities. Syväjärvi et. al. (2009) had aimed to the direction of data mining and information and communication technology (ICT). Ahmadvand et. al. (2010) applied a data mining framework on the database of Tehran municipality. Andrienko et. Al. (1999) proposed to combine applications of techniques of knowledge discovery in databases (KDD) with various methods of interactive classification of spatial objects supported by map displays. Seth and Thill's (2008) analysis of New York City calls into question some assumptions about the functional form of spatial relationships that underlie many modelling and statistical techniques. Solomon et. al. (2006) demonstrated how the use of data mining analysis can be used to evaluate how well cameras that monitor red-light-signal controlled intersections improve traffic safety by reducing fatalities. Minseok and Wil (2008) presented new process mining techniques but also use existing techniques in an innovative manner. Their approach had been implemented in the context of the ProM framework and has been applied in various case studies. They demonstrated the applicability of their techniques by analyzing the logs of a municipality in the Netherlands. Guoqing et. al. (2009) presented a research effort undertaken to explore the applicability of data mining and knowledge discovery (DMKD) in combination with Geographic Information System (GIS) technology to pavement management to better decide maintenance strategies, set rehabilitation priorities, and make investment decisions.

3. Service Oriented Architecture (SOA)

The base of SOA is partitioning the application layer into services and using them on servers. In this study five services are created to supply SOA based data mining solutions for local municipalities. The main reasons that make us choose SOA are; reusability, interoperability, scalability, flexibility and low cost. Web services are created generically. That means all services can work with all kind of dataset.

Data Preparation Web Service (DPWS): This web service is created to prepare data that comes from OMIS for data mining algorithms. Includes 4 child processes. These are; Data Integration, Data Reduction, Data Preprocessing and Data Transformation.

Association Rule Mining Web Service (ARMWS): This service is used to discover hidden relationships from large datasets. FP-Growth (Han et al., 2000) algorithm is used in this service.

Outlier Detection Analysis Web Service (ODAWS): Outliers are detected via this service. Local Outlier Factor (LOF) (Breunig et al., 2000) algorithm is used.

Classification Web Service (CWS): Classification operations are performed through this web service. Bayes (Thomas Bayes, 18th Century) and C4.5 (Quinlan, 1993) algorithms are used.

Clustering Web Service (ClustWS): This web service is used for clustering. KMeans++ (Arthur and Vassilvitskii, 2007) algorithm is used.

4. System architecture

Data mining solutions for Local Municipalities works integrated with OLGU Management Information System (OMIS (OYBS-MIS)), and it is called OMIS-DM(OYBS-DM) as shown in Figure 1.

- Data from OMIS are sent to the Data Preparation Web Service
- After data preparation operation, data that is ready for data mining, is collected in the Data Warehouse
- This data is used in data mining web services according to scenarios.
- Results of data mining are saved in database server.

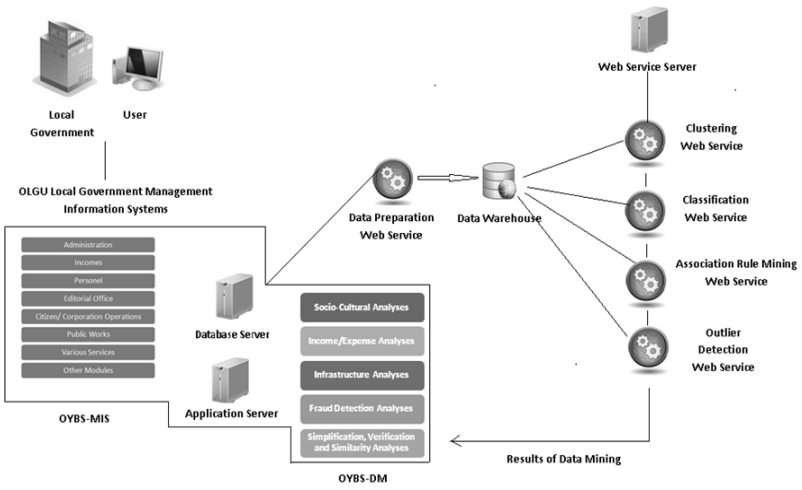

Figure 1: Deployment diagram

5. Scenarios

5.1 Socio-cultural analyses

These kinds of analyses are created to analyze citizens, corporate foundations like bank, company, cooperative and municipal employee.

Employee Analyzing

The purpose of this scenario is to cluster local government employees according to number of their children, age, duration of working, the number of days that employee didn't work and education status.

This scenario is important to make common decisions about municipal employees. For example municipal or department manager may want to warn employees who have big number of non working days by sending mails to this cluster or manager may want to make advance payments to employees whose duration of working bigger than 10 years and number of children bigger than 3 but have small non working days. In this situation manager is able to find this cluster easily. Also correlation between employee's attributes can be found via this scenario.

Clustering service is used for this scenario.

Citizen Analyzing

Citizens can be profiled according to their demographic properties, so local management decisions can be made suitable to these profiles.

These profiles can be useful for election campaigns, promotions and charity events. For example according to our sample analyses 56,8% of Karşıyaka (county of Izmir, Turkey) citizens are married, ages are between 65 – 80 and female. In this situation we can see that promotions which focussed on youths should not be done in this area. Or this area is not suitable for education charity events. If the mayor wants to perform election campaign then mayor should create strategies for majority to make satisfy citizens. For example new hiking trails may be more suitable than new tennis courts for these citizens.

This scenario uses Association Rule Mining service.

Staff Analyzing

Local government staff can be profiled according to their demographic properties, so local management decisions can be made suitable to these profiles.
Association Rule Mining service is used for this scenario.

Corporate Foundation Analyzing

The aim of this scenario is to cluster corporate foundations according to their annual payments. This scenario can be useful to find corporate foundations which pay duties regularly. For example if municipal wants to serve some kind of services to the foundations then municipal may choose foundations which pay duties regularly, to reward them. Or corporate foundation cluster which do not pay duties regularly can be find easily by using this scenario and they can be warned.
This scenario uses Clustering service.

5.2 Income/expense analyses

These analyses include payment of duty like real estate, environmental and expense of moveable material logs analyses.

Moveable Material Analyzing

The aim of this scenario is to find material-year-supplier-amount relations. This scenario may be used to facilitate moveable material input process. For example according to analyze, if 53% of logs 150.01.04.01 and 150.01.01.03 moveable codes are used together then, if user enters 150.01.04.01 code then 150.01.01.03 code will be showed on the screen automatically. This scenario will facilitate and accelerate input operations.

Association Rule Mining service is used by this scenario.

Estimation of Wages

User can estimate wages of employees by analyzing previous employee records through this scenario. To make prediction, age of employees, duration of working as years, education level, job and wage are used. This scenario is useful to provide fairness. Because there will be concrete assessment criteria to determine wages.

Classification service is used for this scenario.

Classifying Citizens According to Real Estate Tax

The main aim of this scenario is to classify citizens according to their real estate tax payments and so tax payments of next term or year can be estimated. According to this estimation some precautions can be supplied.

This scenario uses classification service.

Distribution of Citizens Delaying Real Estate Tax

The purpose of this scenario is to find distribution of the citizens who delay the real estate tax payment, on the map. Sample analyze result is shown in Figure 2. According to this figure it can be seen that green pin symbolizes citizens who pay the highest taxes. If we scan the map in Figure 2, then we can see that green pin is at Mavişehir District.

This scenario uses Clustering service.

Income Operations Analyzing

The main aim of this scenario is to explore which types (MS, BİL, A, D, ...) of which income operations (real estate, water, environmental cleaning etc.) are done frequently by which users on which months. This scenario is useful to prevent wrong operations. For example according to analyze 27% of rules are User Name=”Ali Yılmaz” and Income Type Code = “Real Estate Tax”. If this user does any

operation which Income Type Code=”Sanitation Tax” then system will warn him about operation. This prevents possible wrong operations.

This scenario uses Association Rule Mining service.

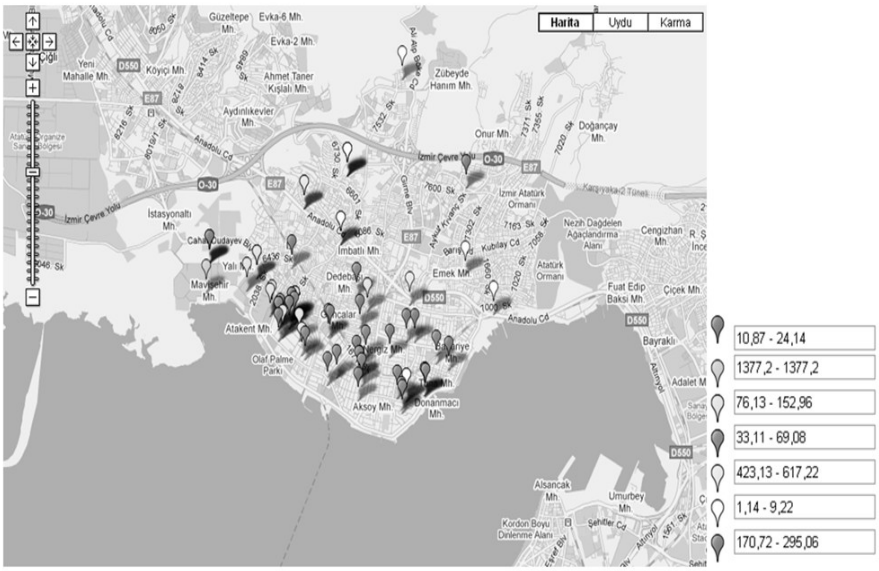

Figure 2: Distribution of citizens

5.3 Infrastructure analyses

These kinds of analyses are created to analyze water and electricity consumptions and distribution of corporate foundation according to their depts.

Water Notice Analyzing

This scenario’s purpose is to cluster citizens according to their water consumption. This scenario can be useful when municipal authority wants to warn citizens who consume water higher than limit.

This scenario uses Clustering service.

Electricity Consumption Analyzing

Aim of this scenario is to detect outliers by analyzing electrical consumptions. Example analyze result is shown at Table 1. As it can be seen from the table, the highest consumption was done in August in 2007. Also, the least consumption was done in May in 2009. Reasons of these situations can be searched according to analysis result.

Table 1: Electricity consumption analyzing

Month	Year	Total
5	2009	35,94414
1	2009	1981,184
12	2009	3620,022
8	2007	4075,097

This scenario uses Outlier Detection service.

Distribution of Corporate Foundation

The purpose of this scenario is to view corporate on the map by clustering them according to their depts. This scenario is similar to “Distribution of Citizens Delaying Real Estate Tax” scenario.

This scenario uses Clustering service.

5.4 Fraud detection analyses

Cancellation operations at cash desks, personnel performances and fuel oil analyses belong to this kind of analyzing.

Logs Analyzing

The main aim of this scenario is to determine who performs which transactions on which computer, on which days on which tables. This scenario can be useful for table update or transportation operations. For example if the minimum support value of "Table Name="Income", Time="08.00-12.00" and Day="Wednesday"" rule is 46% then we can realize that on Wednesday at 08.00-12.00 critical operations should not be applied on "Income" table, this can cause a big problem. Because generally personnel work on this table on Wednesday between 08.00 and 12.00.

Possible abnormalities can be caught by using pre-created general user profile information.

This scenario uses Association Rule Mining service.

Fuel Oil Analyzing

Outliers can be detected by analyzing fuel oil consumptions through this scenario. Benefit of this scenario is similar to "Electricity Consumption Analyzing" scenario. Table 2 shows sample Fuel Oil Analyzing result.

Table 2: Fuel oil analyzing

Month	Year	Total (TL)
9	2008	87976,56
7	2009	95000,128
12	2009	98520,022

This scenario uses Outlier Detection service.

Cash Desk Analyzing

The goal of this scenario is to analyze cash desk operations that are performed by cashier and to detect outliers. Table 3 shows sample result of analysis. As shown below, the collector whose ID is 13 deletes records that belong to Ahmet Kaleli 73 times. This is not normal. This situation should be asked collector.

Table 3: Cash desk analyzing

Collector ID	Number of Deletion	Name	Surname
13	73	AHMET	KALELİ

This scenario uses Outlier Detection service.

5.5 Simplification, verification and similarity analyses

These analyses include User Account and Accountancy Analyses.

User Account Analyzing

The purpose of this scenario is to profile user accounts. According to this scenario answer of which user generally when studies question can be found. For example "User Name="Ali Yıldırım" Start Time="08.00-12.00" Session Duration="4 Hours"" rule is found with 57% minimum support. In this situation for example we can conclude that extra works should not be demanded from "Ali Yıldırım" between 08.00 and 12.00. Because he works dense at these hours.

This scenario uses Association Rule Mining service.

Accountancy Analyzing

The main aim of this scenario is to detect which accounts are entered together. So, accounts will be viewed on the screen automatically. This provides easy information entrance.
This scenario uses Association Rule Mining service.

6. Application

OMIS-DM was developed by using Visual Studio .Net 2010 platform. Silverlight technology is used for interfaces.

Local government personnel can use OMIS-DM without having any theoretic information about data mining via user friendly interfaces. User can do analyzes just by pressing a button. User can save analysis results or print results as a report.

User can add new scenarios by using “new scenarios” screen by defining queries, data preparation operations and data mining technique.

All data preparation operations are done in the background automatically. For example if a scenario uses clustering web service then normalization operation is done without showing on the screen.

Classification Module

General view of classification screens are shown in Figure 3. Results are shown in grid view and tree view. User can do estimation operation at the bottom of the screen.

Classification screens are divided into 3 parts. First part includes grid view. All details of results are shown at this part. If-then rules that are acquired from decision tree are listed with proportion details in grid view. Second part is allocated to show decision tree by using tree view. For example; according to example that is shown in Figure 3, if age>80 then probability of timely real estate tax payment is ≤20. User can see this result by looking at the tree view.

Third part is created to make prediction. User enters values of necessary attributes, and then makes prediction by pressing only one button. If user wants to classify more than one record, then user loads related file by using other button.

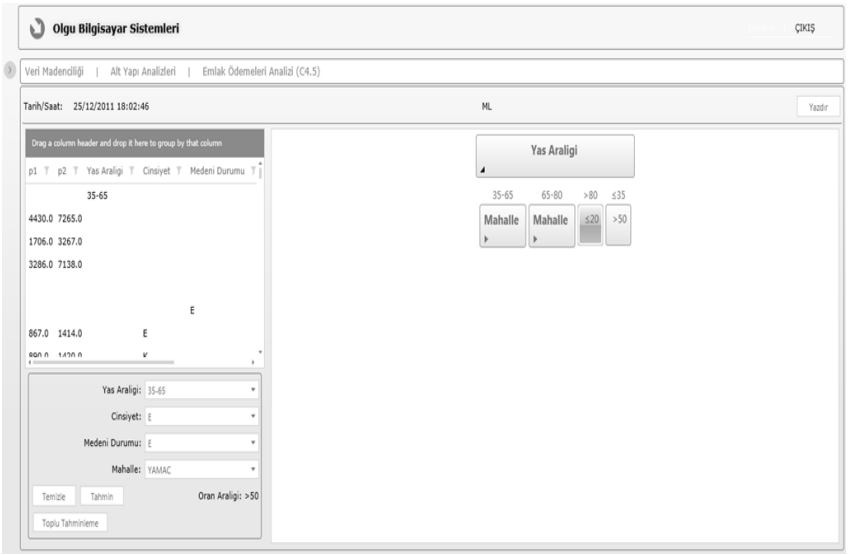

Figure 3: Classification screen

Clustering Module

Clustering results are shown on the map. Each colour symbolizes different clusters. Firstly, records are clustered, and then, they are shown on the map according to their addresses as shown in Figure 4.

User can make comment about citizens just by looking at the map. In the example shown in Figure 4, citizens are clustered according to their real estate payments by using “Distribution of Citizens Delaying Real Estate Tax” scenario. For example while purple and pink pins are intensive at the coastline, blue pins are intensive at upcountry. So, reasons of that situation can be searched and suitable solutions can be found.

Detailed results are shown in grid view as well.

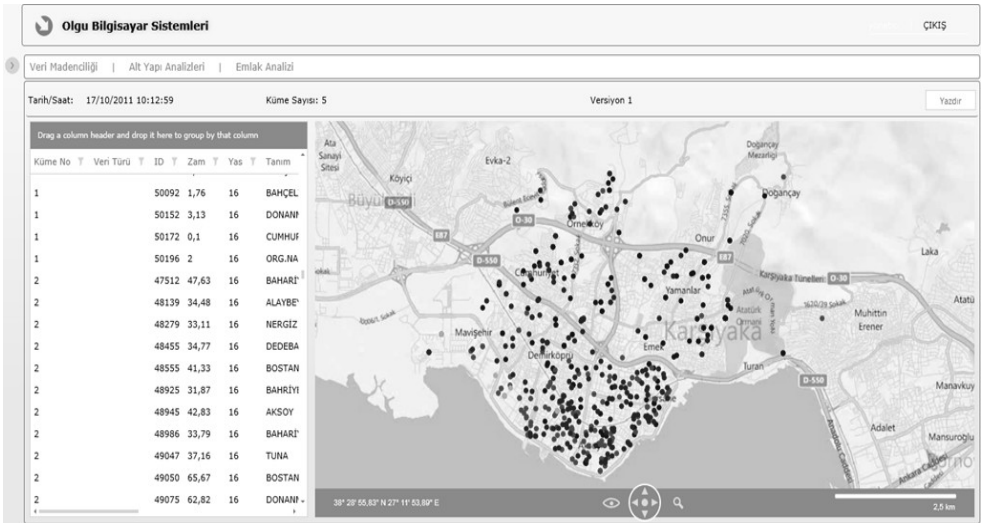

Figure 4: Clustering screen

Outlier Detection Module

Results of scenarios that use Outlier Detection service are shown on the screen like in Figure 5. In this example Electricity Consumption Analyzing scenario results are shown in Figure 5. Outlier value of outliers is “1”, while others are “0”. Results are shown with month and year details. As shown in Figure 5, there are outliers in electricity consumption on October 2009, and on December 2007 for related local government. While there is so much electricity consumption on October 2009, this consumption is so small on December 2007.

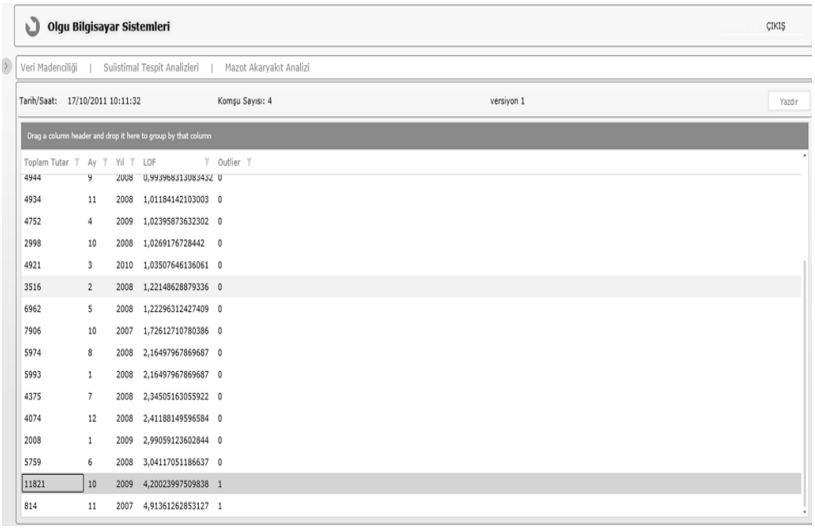

Figure 5: Outlier detection screen

Association Rule Mining Module

The results of scenarios that use Association Rule Mining service are shown with a grid view and a bar chart as in Figure 6.

Association Rule Mining screens are divided into 2 parts. First part includes grid view. Relations are shown with support and number values. There is a bar chart in the second part of the screen. This chart shows relations between items. In the example that is shown in Figure 6, "600A" account code was selected at the top of the graphic. And then account codes that work with 600A are shown at the chart with ratio values.

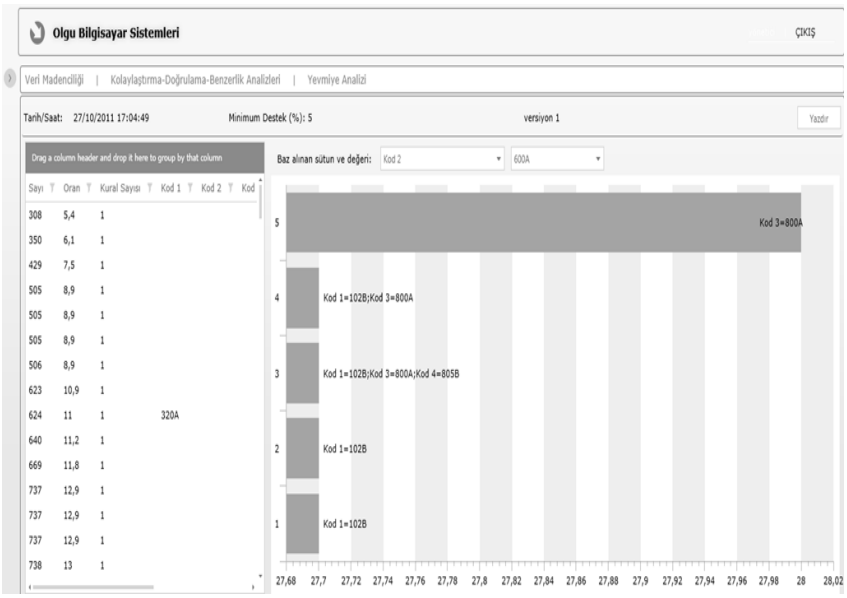

Figure 6: Association rule mining screen

7. Practice

OLGU Computer Systems works with more than 300 local municipalities all over Turkey. The aim is to supply usage of this software by all of our local municipalities. Initially, three big local municipalities were assigned as pilot municipalities to this software OMIS-DM. These municipalities were; Urla Municipality, Eskişehir Metropolitan Municipality and Karşıyaka Municipality. Software was deployed to servers of these municipalities in August in 2012. Training staff about data mining, analyses, interpreting graphics and results took two weeks. After training process, authorized employee started to analyze operation and use scenarios.

As we predicted, authorized employees could perform analyses easily just pressing one button. But at first interpreting results were not as easy as performing analyses. Clustering, classification and outlier detection analyses were understandable for them. Because for example results of clustering analyses like "Distribution of Citizens" are shown on map. So interpreting results is easy for employees. But association rule mining analyses were complex and hard to understand and interpret analyses' results. Another reason is that classification, clustering and fraud detection terms are used in daily life. But association rule mining is new for them. And also graphs of association rule mining analyses are more complex than other kind of analyses.

According to the last state, authorized employees can perform all scenarios and interpret results of analyses easily. Most commonly used scenarios are "*User Account Analyzing", "Electricity Consumption Analyzing" and "Distribution of Citizens Delaying Real Estate Tax"* in local municipality. Also some private companies want profile of citizens which are gotten from "Citizen Analyzing" for their advertising campaign. In a pilot municipality, 2 important frauds were detected. And also employees start to pay attention to their operations on computer and login and logout times by force of "*User Account Analyzing"* and "*Logs Analyzing"* scenarios.

Nowadays, employees try to achieve to add new scenarios. But this operation can be achieved by authorized employees who have data mining algorithms and SQL (Structured Query Language) knowledge.

8. Conclusion

Data mining solutions for local municipalities provide that local municipalities can discover hidden patterns, relationships, changes, irregularities and rules from large datasets. This study proposes a system called OMIS-DM that covers socio-cultural analyses, income/expense analyses, infrastructure analyses, fraud detection analyses and simplification, verification and similarity analyses based on SOA.

Seventeen scenarios were created through municipal employees' various requirements. All scenarios were designed to provide easy usage and detailed reporting. Municipal staff can perform analyses by using only one button and interpret analyses' results easily. Various graphics were used at user interface to provide easy reporting.

Local governments can perform their decision making process more rational, more accurate and faster through OMIS-DM and local government personnel can discover knowledge without having any theoretic information about data mining, by using developed business intelligence software OMIS-DM.

Results of pilot practices show that this intelligent system is usable at local municipality after short training process and also this system provides useful improvements in local municipalities. According to results of two-month trying process in pilot municipalities, performance of employees increases by force of scenarios like "*User Account Analyzing"* and "*Logs Analyzing"*, in addition input operations are facilitated and accelerated by "*Moveable Material Analyzing".* Some possible wrong operations are prevented through "*Income Operations Analyzing"*.

Finally, we foresee that after a while, authorized employee can add new scenarios according to requirement of the municipality in addition to performing analyses.

Acknowledgements

Special thanks are directed to TUBITAK (TSTRI (Turkish Scientific and Technical Research Institute) for its financial support throughout project.

References

Aggarwal, C., Sun, Z. and Yu P. (2002) 'Fast Algorithms for Online Generation of Profile Association Rules'*, IEEE Trans. Knowl. Data Eng*, vol. 14, no. 5, pp, 1017-1028.

Ahmadvand, A.M., Bidgoli, B.M. and Akhondzadeh, E. (2010) 'A Hybrid Data Mining Model for Effective Citizen Relationship Management: A Case Study on Tehran Municipality', *e-Education, e-Business, e-Management, and e-Learning*, 2010 IC4E '10. International Conference on, Sanya.

Andrienko, G.L. and Andrienko, N.V. (1999) 'Data mining with C4.5 and interactive cartographic visualization'*, User Interfaces to Data Intensive Systems, 1999. Proceedings.*

Arthur, D. and Vassilvitskii, S. (2007) 'K-means++ : the advantages of careful seeding', *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027-1035

Bayes, T. (18th Century) 'An Essay Towards Solving a Problem in the Doctrine of Chances'.

Breunig, M.M., Kriegel, H., Ng, R.T. and Sander, J. (2000) 'LOF: Identifying Density-based Local Outliers'*, ACM SIGMOD Record 29:93.*

Han, J., Pei, H. and Yin, Y. (2000) 'Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data'*, (SIGMOD'00, Dallas, TX), ACM Press, New York, NY, USA.*

OLGU Computer Systems (2009) http://www.olgu.com.tr/

Poles, S. and Margonari, M. (2009) 'A modeFRONTIER Application: Data Mining the Italian Municipalities', *Newletter EnginSoft*, vol. 6, no. 2, pp. 31-37.

Solomon, S., Nguyen, H., Liebowitz, J. and Agresti W. (2006) 'Using data mining to improve traffic safety programs', *Industrial Management & Data Systems*, vol. 106, no. 5, pp. 621-643.

Song, M. and van der Aalst, W. M.P. (2008) 'Towards comprehensive support for organizational mining', Elsevier Decision Support Systems, vol. 46, no. 1, pp. 300-317.

Spielman, S.E. and Thill, J. (2007) 'Social area analysis, data mining and GIS', *Elsevier Computers, Environment and Urban Systems*, vol. 32, no. 2, pp. 110 – 122.

Syväjärvi, A., Stenvall, J., Laitinen, I. and Harisalo, R. (2009) 'Information Management as Function of Data Mining and ICT in City Government', *EGPA 2009, Malta.*

Quinlan, J.R. (1996) *C 4.5: Programs for Machine Learning,* Morgan Kaufmann Publishers.

Zhou, G., Wang, L., Wang, D. and Reichle, S. (2010) 'Integration of GIS and Data Mining Technology to Enhance the Pavement Management Decision Making', *Journal of Transportation Engineering*, vol. 136, no. 4, pp. 332-341.