# DATA MINING IN MEDICAL RECORDS FOR THE ENHANCEMENT OF STRATEGIC DECISIONS: A CASE STUDY

Yilmaz GOKSEN[1], Mete EMINAGAOGLU[2] and Onur DOGAN[3]

[1] Dokuz Eylul University, IIBF Dokuzcesmeler Buca Izmir-Turkey, yilmaz.goksen@deu.edu.tr
[2] Yasar University, Selcuk Yasar Kampusu Agacli Yol Bornova Izmir-Turkey, mete.eminagaoglu@yasar.edu.tr
[3] Dokuz Eylul University, IIBF Dokuzcesmeler Buca Izmir-Turkey, onur.dogan@deu.edu.tr

***Abstract:*** *The impact and popularity of competition concept has been increasing in the last decades and this concept has escalated the importance of giving right decision for organizations. Decision makers have encountered the fact of using proper scientific methods instead of using intuitive and emotional choices in decision making process. In this context, many decision support models and relevant systems are still being developed in order to assist the strategic management mechanisms. There is also a critical need for automated approaches for effective and efficient utilization of massive amount of data to support corporate and individuals in strategic planning and decision-making. Data mining techniques have been used to uncover hidden patterns and relations, to summarize the data in novel ways that are both understandable and useful to the executives and also to predict future trends and behaviors in business. There has been a large body of research and practice focusing on different data mining techniques and methodologies.*

*In this study, a large volume of record set extracted from an outpatient clinic's medical database is used to apply data mining techniques. In the first phase of the study, the raw data in the record set are collected, preprocessed, cleaned up and eventually transformed into a suitable format for data mining. In the second phase, some of the association rule algorithms are applied to the data set in order to uncover rules for quantifying the relationship between some of the attributes in the medical records. The results are observed and comparative analysis of the observed results among different association algorithms is made. The results showed us that some critical and reasonable relations exist in the outpatient clinic operations of the hospital which could aid the hospital management to change and improve their managerial strategies regarding the quality of services given to outpatients.*

***Keywords:*** Decision Making, Medical Records, Data Mining, Association Rules, Outpatient Clinic.

**JEL Classification Codes : C80, I11, D83**

## 1. INTRODUCTION

In today's competitive world, information and knowledge has become the absolute power for both establishing and managing organizations in a successful and resilient manner. The key players in global economy, managers and strategic decision makers namely, strive for better and more reliable decision making / decision support systems and mechanisms with the aid of continually improving information technologies and automated business intelligence models.

Since the 1990s, the socio-economic concept has generally been referred to as the "information and knowledge society". The profound changes that have occurred in methods of production and in economic relations are accepted to increase the importance of the exchange of intangible goods, consisting for the most part of transfers of information. The acceleration in the pace of current transformation processes is shown to be due to two factors where the first one is globalization and the second one is the impact of new information technologies regarding the massive spread of the Internet and mobile devices. The spreading use of low-cost massive data storage technologies and the wide availability of Internet connections have made very large amounts of data available to organizations, governments and people. The enterprises that are capable of transforming data into information and knowledge can use them to make quicker and more effective decisions and thus to achieve a competitive advantage (Vercellis, 2009).

Consequently, the rapid growth and integration of information technologies, digital networks, software and database systems and the availability of massive amount of electronic data provide people with a vast new resource that can be analyzed to optimize industrial systems, uncover financially valuable patterns, minimize investment risks, make successful strategic decisions, and so on (Costea & Eklund, 2009; Changchien, et al., 2004). Even, some interesting studies have been applied in fraud detection in several business area including medical records and systems (Ortega, et al., 2006). To undertake these large data analysis projects, researchers and practitioners have adopted established algorithms from statistics, machine learning, neural networks, and databases and have also developed new methods targeted at large data mining problems (Hand, et al., 2001; Zhang & Zhou, 2004). Data mining can be defined as the extraction of useful information from large data sets or databases. It is globally accepted as a new discipline, lying at the intersection of statistics, machine learning, data management and databases, pattern recognition, artificial intelligence, and other areas. All of these are shown to be related with certain aspects of data analysis, so they have much in common—but each also has its own distinct flavor, emphasizing particular problems and types of solution (Hand, et al., 2001; Rokach & Maimon, 2008).

With the advent of computers and the information age, statistical problems have exploded both in size and complexity. Challenges in the areas of data storage, organization and searching have led to the new field of "data mining"; statistical and computational problems in biology and medicine have created "bioinformatics." Vast amounts of data are being generated in many fields, and the statistician's job is to make sense of it all: to extract important patterns and trends, and understand "what the data says." This is generally described as "learning from data". In other words, data mining refers to the search of large, high-dimensional, multi-type data sets, especially those with elaborate dependence structures or patterns where the search for valuable structure or patterns is based on statistical methodologies (Hastie, et al., 2009).

Data mining is also accepted as a stage of a larger process known as "Knowledge Discovery in Databases" (KDD). Knowledge discovery is defined as a process, in several stages, not trivial, interactive and iterative, for identification of new valid understandable and potentially useful patterns from large data sets. Thus, the use of data mining is intended to support the discovery of patterns in databases in order to transform information in knowledge, to assist the decision making process or to explain and justify it. Data mining can be defined as an automatic or semiautomatic patterns discovery in great amounts of data, where these patterns can be perceived as useful (Nedjah, et al., 2009).

There are several different data mining models, methodologies, tools, algorithms and implementations, however, in most cases, the stages or phases of a data mining process is defined and grouped generally based on CRISP–DM standard (Larose, 2005). These phases are described as follows (Larose, 2005):

1. Business understanding phase
2. Data understanding phase
3. Data preparation phase
4. Modeling phase
5. Evaluation phase
6. Deployment phase

It is usually noted that the most crucial phases is the first three stages where a good understanding of the business, data structure, the scope and objectives of the study and purifying or cleaning the data (preparation phase) is the key factors for a successful data mining implementation (Dasu & Johnson, 2003). In today's data mining models and relevant technologies, most of them are aimed at one or more of the following common objectives or tasks, which also gives us the summary of primary goals and alternative approaches of data mining concept (Witten, et al., 2011; Larose, 2005):

- Description
- Estimation
- Prediction
- Classification
- Clustering
- Association

Since in this study, association is aimed and implemented, it will be shortly described in this paper. The association task for data mining is the job of finding which attributes "go together." Most prevalent in the business world, where it is known as affinity analysis or market basket analysis, the task of association seeks to uncover rules for quantifying the relationship between two or more attributes. Association rules are of the form "If antecedent, then consequent," together with a measure of the support and confidence associated with the rule. Examples of association tasks in business and research include examining the proportion of children whose parents read to them who are themselves good readers, or predicting degradation in telecommunications networks, finding out which items in a supermarket are purchased together and which items are never purchased together, and so on (Larose, 2005; Hastie, et al., 2009). Association rules can "predict" any of the attributes, not just a specified class, and can even predict more than one thing (Witten, et al., 2011).

In this study, a data mining model and a proper data mining implementation was achieved in a outpatient clinic database system. The executive board of the hospital was especially concerned about improving the quality of the service given to patients in the outpatient clinic, as well as improving the work conditions for the medical staff and increasing the efficiency and throughput of the business processes. They had a large database that stores the incoming patients' records to the outpatient clinic, however the previous reports generated from the database system hadn't provided them with the type of information or hints / clues that can increase their knowledge in their case. Since, their primary concern and requirement was to extract or find out some hidden valuable relations in the data that can guide them in making decisions for change management in their daily operations; the association task was chosen within the data mining implementation. It is aimed to find some accurate and meaningful specific conditions and criteria that can predict certain situations, conditions or results that can be set as rules for hospital's strategic decisions by using the record sets and fields (attributes) in the database. It should also be mentioned that some previous studies about data mining applications amongst medical and healthcare records have been developed successfully by other researchers (Riha, et al., 2002; Silver, et al., 2001; Morik, et al., 1999). The implementation of our study, all necessary phases of data mining process including data collection, preparation, modeling, execution of algorithms and test, observations of the results and discussions are given in the following sections of this paper. □

## 2. CASE STUDY

The implementation of this research was carried out in one of the medium-sized public hospitals in Izmir, where Izmir is known to be the third biggest (population size and socio-economic parameters) city in Turkey. Due to the privacy and legal concerns of hospital managers, the name of the hospital is not explicitly given in this study. The total number of medical staff working in the hospital is around 580, the inpatient bed capacity is given as 300 and it has three different outpatient clinic buildings and two main buildings. During the interviews conducted with the hospital's executive members in the business analysis of this study, it was observed that senior management is becoming more involved in developing quality assurance standards, improving their IT infrastructure and adapting their major business process to technological advances.

### 2.1 Data Collection and Preparation

The data were collected from the hospital's outpatient clinic Oracle 9i v.9.2.0.1 database system. It was made up of a four-month period of data that were recorded to the system whenever the patients arrived at the clinic. In the original data, there were 21 different attributes (fields) for each record and a total of 257668 records. However, before the data analysis and mining process, some of these fields were eliminated and were not used in the study due to the irrelevance of these fields for the aim of this study. Also, some of the records had null data (wrong entries by data operators in the hospital) and these records were also discarded from the data set. By this way, the data cleaning and preparation process, which is an important step in data mining was carried out. As a result, a total of 256816 records and 9 different fields were collected and they were used for data mining analysis in the study. The descriptive statistics of the whole data set denoted in Figure 1 and the name and properties of the fields are given in Table 1.

It should be noted that since this data is collected from a Turkish hospital system, some of the records in the original data set are written and stored in Turkish language but some of them are translated into English in this study, whenever necessary. A sample screen shot of the record set is also given Figure 2.

In the data set, 146801 were the records of female patients (coded as "K" in the database records) which is 57.16% and the remaining 110015 records were belonging to male patients (coded as "E" in the database records), which is 42.84%. The day of the week field stores nominal values representing any day of the week days (Sunday, Monday, etc.) where it corresponds to the "Date" field in the same record. Since, the data was collected in a period of four months, a total of 120 distinct field values were collected in the "Date" field. The "Time" field denotes the hour and minute value of the time (24-hour format) which the patient's record is entered into the hospital system. The "2-hour Period" field is a nominal label which denotes the time period that corresponds to its time value. For instance, if a record has a "Time" value between 00:00 and 02:00, then its "2-hour Period" will be labeled as "0". In the same manner, if the time of the record is between 06:00 and 08:00, then its "2-hour Period" will be set as "3". Hence, this "2-hour Period" has distinct 12 alternative values ranging from 0 to 11.

**Table 1. The name of the fields in the whole data set**

| FIELD NAME | SAMPLE VALUES |
|---|---|
| **Patient Gender** | **Male, Female** |
| **Day of Week** | **Sunday, Monday, etc.** |
| **Date** | **mm/dd/yyyy** |
| **Time** | **hh:mm** |
| **2-Hour Period** | **0, 1, 2,….11, 12** |
| **Department Code** | **100110, 200211, etc.** |
| **Patient Type Index** | **1, 2, 3,……………56, 57** |
| **Diagnosis Code** | **N21.1, S65.3, G46.8-J20.9, etc.** |
| **Case Explanation** | ***(Summary Of Each Record Explaining The Patient And The Incident, Written In Turkish)*** |

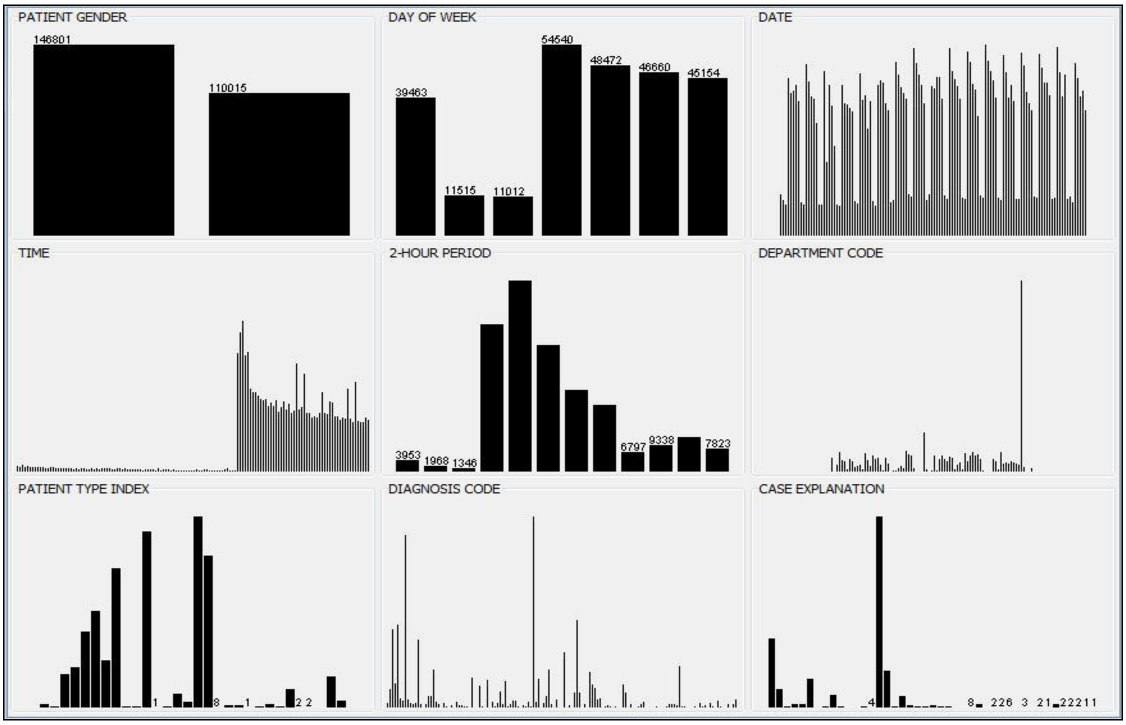


**Figure 1. The descriptive statistics of the whole data set in the study**

| PATIENT GENDER | DAY OF WEEK | DATE | TIME | 2-HOUR PERIOD | DEPARTMENT CODE | PATIENT TYPE INDEX | DIAGNOSIS CODE | CASE EXPLANATION |
|---|---|---|---|---|---|---|---|---|
| E | Thursday | 2/18/2010 | 18:21 | 9 | 400610 | 9 | T14.8-T14.9 | 'PANSUMAN BİRİMİ VAKA DIŞIDIR' |
| K | Tuesday | 4/27/2010 | 09:53 | 4 | 400412 | 7 | F32.9 | 'VAKA BAŞI HASTASIDIR' |
| K | Thursday | 2/18/2010 | 20:44 | 10 | 400710 | 5 | R51 | 'YEŞİLKART HASTA TÜRÜ VAKA BASI UYGULAMASINA DAHİL DEĞİLDİR' |
| E | Wednesday | 3/31/2010 | 07:59 | 3 | 200311 | 9 | T15.1 | 'GÜNÜBİRLİK HASTASI((PAKET KARŞILIĞI TETKİĞİ YOK))' |
| K | Wednesday | 3/17/2010 | 07:49 | 3 | 100610 | 9 | R42 | 'AYNI GÜN BAŞKA VAKA BAŞI GELİŞİ VAR' |
| K | Saturday | 3/13/2010 | 09:44 | 4 | 101410 | 22 | Y60.3 | 'GÜNÜBİRLİK HASTASI((PAKET KARŞILIĞI TETKİĞİ YOK))' |
| K | Monday | 3/22/2010 | 12:46 | 6 | 100626 | 13 | E11.9 | 'VAKA BAŞI HASTASIDIR' |
| K | Wednesday | 1/27/2010 | 07:23 | 3 | 200620 | 23 | N83.2 | 'ON GÜN HASTASI EK/10C+GENETİK TETKİKİ YOK (01.04.2009 sonrası)' |
| E | Wednesday | 4/28/2010 | 07:59 | 3 | 200401 | 9 | L05.9 | 'ON GÜN HASTASI EK/10C+GENETİK TETKİKİ YOK (01.04.2009 sonrası)' |
| E | Tuesday | 4/27/2010 | 12:50 | 6 | 100316 | 13 | J45.8 | 'VAKA BAŞI HASTASIDIR' |
| E | Wednesday | 4/14/2010 | 21:00 | 10 | 400710 | 9 | M79.1-M79.1/MIYALJI | ' ACİL POLK. BİRİMİ VAKA DIŞIDIR' |
| E | Monday | 3/15/2010 | 14:12 | 7 | 100627 | 22 | K29.7 | 'VAKA BAŞI HASTASIDIR' |
| K | Saturday | 4/24/2010 | 20:45 | 10 | 400710 | 22 | N30.9 | ' ACİL POLK. BİRİMİ VAKA DIŞIDIR' |
| K | Monday | 3/15/2010 | 07:33 | 3 | 200310 | 5 | H25.1 | 'YEŞİLKART HASTA TÜRÜ VAKA BASI UYGULAMASINA DAHİL DEĞİLDİR' |
| K | Monday | 1/25/2010 | 11:03 | 5 | 100317 | 13 | K52.8 | 'VAKA BAŞI HASTASIDIR' |
| K | Thursday | 1/14/2010 | 10:50 | 5 | 101110 | 22 | M67.4 | 'VAKA BAŞI HASTASIDIR' |
| E | Monday | 4/19/2010 | 10:06 | 5 | 200312 | 8 | H10.8 | 'ÜCRETLİ HASTA' |
| K | Tuesday | 4/13/2010 | 07:50 | 3 | 200710 | 13 | 'I25.9 -I25.9' | 'VAKA BAŞI HASTASIDIR' |
| K | Monday | 2/22/2010 | 14:09 | 7 | 201010 | 4 | 'R07.3 -R07.3' | 'VAKA BAŞI HASTASIDIR' |
| K | Monday | 3/22/2010 | 12:40 | 6 | 101410 | 22 | Y60.3 | 'ENJEKSİYON POLK. BİRİMİ VAKA DIŞIDIR' |
| E | Thursday | 4/29/2010 | 01:38 | 0 | 400710 | 20 | Z02.9 | 'CUMHURİYET BAŞ SAV. HASTA TÜRÜ VAKA BASI UYGULAMASINA DAHİL DEĞİLDİR' |
| E | Monday | 3/15/2010 | 07:41 | 3 | 100610 | 9 | E78.4-E11.5 | 'VAKA BAŞI HASTASIDIR' |
| K | Monday | 4/19/2010 | 07:36 | 3 | 101111 | 13 | M79.8 | 'VAKA BAŞI HASTASIDIR' |
| E | Thursday | 1/21/2010 | 10:55 | 5 | 400710 | 9 | M79.1-M79.1/MIYALJI | ' ACİL POLK. BİRİMİ VAKA DIŞIDIR' |
| E | Friday | 4/30/2010 | 07:34 | 3 | 100610 | 23 | E11.9 | 'VAKA BAŞI HASTASIDIR' |
| K | Tuesday | 3/30/2010 | 09:24 | 4 | 200210 | 5 | J44.9 | 'YEŞİLKART HASTA TÜRÜ VAKA BASI UYGULAMASINA DAHİL DEĞİLDİR' |
| K | Tuesday | 4/27/2010 | 07:22 | 3 | 200210 | 13 | 'I25.9 -R07.3' | 'VAKA BAŞI HASTASIDIR' |
| K | Thursday | 2/25/2010 | 06:59 | 3 | 100616 | 13 | I10 | 'VAKA BAŞI HASTASIDIR' |
| E | Friday | 2/26/2010 | 08:01 | 4 | 200311 | 23 | H11.0-H11.0/PTERJIYUM | 'GÜNÜBİRLİK HASTASI(PAKET KARŞILIĞI TETKİĞİ VAR)' |
| K | Sunday | 3/28/2010 | 11:35 | 5 | 400610 | 5 | T14.8-T14.9 | 'YEŞİLKART HASTA TÜRÜ VAKA BASI UYGULAMASINA DAHİL DEĞİLDİR' |
| K | Monday | 1/25/2010 | 15:38 | 7 | 100713 | 7 | F32.9-M17.9 | 'AYNI GÜN BAŞKA VAKA BAŞI GELİŞİ VAR' |

**Figure 2. An excerpt from the data set in the study**

The “Department Code” field stores the data that shows where the patient is medically operated in the outpatient clinic. These codes are specifically defined in the hospital and their corresponding department names are known. In the data set, there were 80 distinct department codes. “Patient Type Index” is also another unique specification code in the hospital system which corresponds to the patient’s socio-economic feature. In other words, each code describes some additional property about the patient; whether he / she is retired and retired from which organization, or actively working, or jobless, whether he / she has an active social security number or not, etc. There are a total of 30 distinct label values in the data set ranging between 1 and 58.

“Diagnosis Code” field holds the specific medical codes that summarize the diagnosis assigned to the patient within each incident. “Case Explanation” field is a nominal data field that the data operators or medical staff enters as short summary notes in a standard and specific format which explains the case for each record of that patient regarding the diagnosis given and other observations. In the data set, there were 43 distinct values among the 257668 records.

## 2.2 Methodology and Implementation

Since the initial data set size was large for the data mining process, before analyzing the data, a statistical sampling methodology was used to derive a smaller sample data set (Witten, et al., 2011; Dasu & Johnson, 2003). The sampling size was chosen according to the following criteria:

Original data set size: 257668
Confidence interval (accepted margin of error): ± 2%
Confidence level: 99%

Within these statistical sampling parameters, the minimum recommended sample size could be calculated as 4081 using the statistical sampling size derivation methods (Witten, et al., 2011; Dasu & Johnson, 2003). Thus, a total of 4100 out of 257668 records was selected as the sample data set size in our study. It should also be noted that, the available data set from the outpatient clinic database in the hospital was limited to the first four month period due to the changes in the technological infrastructure of the relevant database systems. However, this sample size and the sample data set can also be considered as a feasible representative sampling amount for a one year period regarding the same statistical confidence level and confidence interval values.

Weka version 3.6.0 was used as the software for the data mining analysis phase in the study. The whole data set was first converted into proper Weka data format and then it was imported into the application. After this step, all of the numerical fields (attributes) were transformed into their nominal values. This was also a crucial step in the analysis since most of the associative data mining algorithms that are supplied within Weka software only works for nominal data (Witten, et al., 2011). Before the initiation of data mining analysis, the final step was the random sample data set selection of 4100 records out of the generic data set. This was established by the preprocessing and filtering tools that were provided within Weka software.

After these steps, different associative data mining algorithms were executed among the data set and their results were observed and comparatively analyzed. Each of these associative algorithms use different data mining techniques to extract association rules (Witten, et al., 2011; Cho, et al., 2002; Khattak, et al., 2010) and they are summarized in Table 2.

**Table 2. Associative data mining algorithms used in the study**

| ALGORITHM NAME | EXPLANATION |
|---|---|
| **Apriori** | **Generates association rules by finding frequent item sets, generating successively longer candidate item sets from shorter ones that are known to be frequent. It iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence.** |
| **Filtered Associator** | **It allows data to be passed through a filter before it reaches an associator. Both the filter and the base associator are options that the user can configure.** |
| **PredictiveApriori** | **It combines confidence and support into a single measure of predictive accuracy and finds the best n association rules in order.** |
| **Tertius** | **It finds rules according to a confirmation measure, seeking rules with multiple conditions like Apriori, but differing in that these conditions are applied OR operation together, instead of AND operation.** |

Among these four algorithms, the only remarkable and accurate results were observed from the PredictiveApriori associator. The other three data mining algorithms either provided no results or some non-relevant / incorrect association rules. Some of the default parameters were also changed in these three algorithms and additional observations were made in order to achieve some accurate and meaningful results but no further accurate results could be obtained. In the following section, the association rules observed by the PredictiveApriori mining algorithm are given and the interpretation and analysis of these rules are discussed.

### 2.3 Results

PredictiveApriori associator algorithm derived 100 different association rules that were ranked and ordered with several accuracy level values. The rule with the highest accuracy had a value of 0.99498 and the one with lowest accuracy was observed as 0.9733. This accuracy term denotes metrics for ranking the association rules by means of confidence, which is the proportion of the examples covered by the premise that are also covered by the consequent ones (Witten, et al., 2011). Among these rules, some of them had one condition or attribute with a resultant condition where some others had two or more combined condition that corresponds to a specific condition. Some of these association rules derived from Weka software's output panel are shown in Figure 3 with the rule conditions and the accuracy values abbreviated as "acc:";

```
PATIENT TYPE INDEX=22 CASE EXPLANATION= ACIL POLK. BIRIMI VAKA DISIDIR ==>
DEPARTMENT CODE=400710 acc:(0.99489)

DAY OF WEEK=Sunday CASE EXPLANATION= ACIL POLK. BIRIMI VAKA DISIDIR ==> DEPARTMENT
CODE=400710 acc:(0.99471)

2-HOUR PERIOD=10 CASE EXPLANATION= ACIL POLK. BIRIMI VAKA DISIDIR ==> DEPARTMENT
CODE=400710 acc:(0.99463)

PATIENT GENDER=K DEPARTMENT CODE=400710 PATIENT TYPE INDEX=22 ==> CASE EXPLANATION=
ACIL POLK. BIRIMI VAKA DISIDIR acc:(0.99471)

DAY OF WEEK=Sunday DIAGNOSIS CODE=Y60.3 ==> DEPARTMENT CODE=101410 acc:(0.99163)
```

**Figure 3. Some of the association rules derived by PredictiveApriori**

As it can be seen from the sample subset of results in Figure 3, the first rule suggests that if "Patient Type Index" is 22 and "Case Explanation" is a non-standard emergency outpatient clinic, then it should be operated in the department with code 400710 (which is recorded in the hospital's system as emergency service department) with a confidence value of 0.99489. Similarly, the last rule in Figure 3 suggests that if any patient comes on Sunday and its diagnosis code is Y60.3 (unexpected incident due to a surgical operation / wrong medical treatment), then its department code is 101410 (injection service department) with a confidence value of 0.99163.

However, all of these 100 rules had to be analyzed in detail by the authors of this study and the hospital senior management since most of these rules, even providing very high confidence values, in fact were not accurate or meaningful. This was also an expected outcome in this study. This is due to the fact that in all types of data mining models and methodologies; it requires significant human interactivity at each stage (Larose, 2005). Continuous quality monitoring, validation and other evaluative measures must be assessed by human analysts (Larose, 2005). Recent researches are focusing in to find improved mining methodologies that can enable change semi-automated techniques to automated techniques (Asghar & Iqbal, 2009).

After carrying out the analysis of the derived rules, some few but strategic and meaningful conclusions were achieved which were confirmed by the senior management in the hospital. These deductions are given as follows;

- On any day of the week, if the patient arrival time period is 22:00-00:00 at night and the patient is recorded as a non-standard emergency outpatient clinic and if the patient is also retired, then it could be a female patient.
- If the patient's arrival time is 00:00-02:00 at night and the day is Saturday or Sunday and the department is emergency service and if it needs an immediate operation for surgery, it is probably a male patient.
- On Sundays, if the patient has an unexpected incident due to a surgical operation / wrong medical treatment, then emergency service within injection operation is required.
- On any day of the week (except weekends), if the patient arrival time period is 10:00-12:00 in the morning and if the patient is male and if it is a retired patient from government, then he is probably to be served in urology department.

## 3. CONCLUSION

The results from the data mining process of the outpatient clinic records showed us that some necessary precautions and some necessary changes in the daily operations might be achieved by the hospital management that could improve the efficiency and quality of services given to outpatients. For instance, in any work day of the week in the morning period, it is better to have sufficient number of doctors and medical staff in the urology department. At the weekends during the midnight hours, it is crucial to have sufficient number of male nurses available in the emergency service. On Sundays, it is necessary to keep some doctors in the clinic that are experienced in specific surgical operations.

On the other hand, in this study it is shown that only a few relations and knowledge-based conclusions could be made to support and enhance the decisions and managerial strategies of the hospital management. This could be due to two reasons. The first reason is the lack of data in the database system in order to extract valuable associations and relations (in this study only nine fields, in other words, nine criteria in the record set could be used for data mining). The second reason is that the association algorithms used for data mining in this study might not be perfectly fit for this case. To overcome such problems and drawbacks in further studies, more fields or attributes could be included in the data sets and some other association algorithms might be tested as well. Also, it should be noted that sometimes data set itself might not be suitable for association tasks in data mining.

It should also be noted that even if the data mining tools, models and algorithms used in similar studies provide results with high levels of accuracy and confidence, those results and derived rules shall require human intervention. In other words, no data mining algorithm or model can assure 100% correctness and accuracy by itself for a pure and robust automated system, hence, every result must also be checked by experts or managers.

## REFERENCES


1. Asghar, S. and Iqbal, K., 2009. Automated Data Mining Techniques: A Critical Literature Review. *Proceedings of 2009 International Conference on Information Management and Engineering*. Kuala Lumpur, Malaysia, pp. 75-79.
2. Changchien, S., W., et al., 2004. On-line personalized sales promotion in electronic commerce. *In Expert Systems with Applications*, Vol. 27, pp. 35-52.
3. Cho, Y., H., et al., 2002. Personalized recommender system based on web usage mining and decision tree induction. *In Expert Systems with Applications*, Vol. 23, pp. 329-342.
4. Costea, A. and Eklund, T., 2009. Financial Performance Analysis of Scandinavian Telecommunications Companies Using Statistical and Neural Networks Techniques. *In Economic Computation and Economic Cybernetics Studies and Research*, Vol. 43, No. 4, pp. 87-104.
5. Dasu, T. and Johnson, T., 2003. *Exploratory Data Mining and Data Cleaning*. John Wiley & Sons Inc., New Jersey, USA.
6. Hand, C., et al., 2001. *Principles of Data Mining*. The MIT Press, London, UK.
7. Hastie, T., et al., 2009. *The Elements of Statistical Learning, Data Mining, Inference and Prediction, 2nd edition.* Springer, New York, USA.
8. Khattak, A. M., et al., 2010. Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka. *In International Journal of Database Theory and Application*, Vol. 3, No. 1, pp. 13-22.
9. Larose, D., T., 2005. *Discovering Knowledge in Data - An Introduction to Data Mining*. John Wiley & Sons Inc., New Jersey, USA.
10. Morik, K., et al., 1999. Combining Statistical Learning with a Knowledge-Based Approach - A Case Study in Intensive Care Monitoring. *ICML '99 Proceedings of the Sixteenth International Conference on Machine Learning.* Bled, Slovenia.
11. Nedjah, N., et al., 2009. *Innovative Applications in Data Mining.* Springer-Verlag, Berlin, Germany.
12. Ortega, P. A., et al., 2006. A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile. *Proceedings of Conference on Data Mining DMIN06.* Las Vegas, USA, pp. 224-231.
13. Riha, A., et al., 2002. Medical guideline as prior knowledge in electronic healthcare record mining. *3rd International Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields.* Bologna, Italy.
14. Rokach, L. and Maimon, O., 2008. *Data Mining with Decision Trees - Theory and Applications.* World Scientific Publishing Co. Pte. Ltd., Singapore.
15. Silver, M, et al., 2001. Case Study: How to Apply Data Mining Techniques in a Healthcare Data Warehouse. *In Journal of Healthcare Information Management,* Vol. 15, No. 2, pp. 155-164.
16. Vercellis, C., 2009. *Business Intelligence; Data Mining and Optimization for Decision Making*. John Wiley & Sons, Ltd., UK.
17. Witten, I., H., et al., 2011. *Data Mining Practical Machine Learning Tools and Techniques, 3rd Edition.* Elsevier Inc., USA.
18. Zhang, D. and Zhou, L., 2004. Discovering golden nuggets: data mining in financial application. *In IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,* Vol. 4, No. 4, pp. 513-522.