کنفرانس بین المللی مدیریت پروژه
International Project Management Conference

Data Mining in Construction's Project Time Management - Kayson Case Study

Shahram Shadrokh (Assistant Professor)
Sharif University of Technology, Shadrokh@sharif.edu

Seyedbehzad Aghdashi (PhD Student)
North Caroline State University, saghdas@ncsu.edu

Abstract

Data mining is an emerging science in next ten years, so there will be lots of efforts to its development in the near future. The goal of this research is to support project management by presenting various informative insights, which would enable the better understanding of the past dynamics and provide grounds for better planning of the future research project programs. We focused on the time management area. We selected a construction company for gathering data. They have a lot of construction projects in progress. That project has got a lot of information that has recorded and entered in computer. They used Primavera Project Planner (P3) as project management software in their company. Using some processes we transformed their data into Clementine software. Beside that we used some data mining techniques like Classification and Association Rules to develop our results. We used CRISP-DM as our standard procedure that future data mining projects in this area would use this paper's information.

Key words: Data Mining, Project Management, CRISP-DM, Primavera, SQL, Large scale construction

1. Introduction

Data mining is the process of discovering meaningful new correlation, patterns and trends by sifting through large amounts of data stored in repositories [1]. It will be one of the most revolutionary developments of the next decade [2]. MIT technology review chose data mining as one of ten emerging technologies that will change the world [3]. This paper is a case study which is intended to use data mining techniques for discovering knowledge from project management data repositories. Data mining is a multidisciplinary area in which several computing paradigms converge: decision trees, rule induction, artificial neural networks, instance base learning, Bayesian learning, logic programming, statistical programming, etc [4]. It uses some tasks like: visualization, statistics, clustering, classification and association rule mining. [5]

Project Management is the application of knowledge, skills, tools and techniques to project activities in order to meet stakeholder's needs and expectations from a project [6]. PMBOK[1] standard considers nine knowledge areas for project management: 1.Integration, 2.Time, 3.Quality, 4.Cost, 5.Scope, 6.Communication, 7.Human Resource Management, 8.Procurnment, 9.Risk [6]. Project time management includes the process required to ensure timely performance of the project. It consists of activity definition, activity sequencing, activity duration estimation, schedule development and time control [8]. Project time management plays a key role in the project

[1] Project Management Body of Knowledge

management and excellence. It controls project completion within the time defined. Controlling activities' status and comparing it with the project main baseline, and calculating activity delays and floats are some of primary tasks in this section of project management.

Usually project management team is responsible for project time management and they use some software packages for managing and controlling it, some examples of project management software packages are Primavera, and Microsoft Project. Nowadays one the most commonly used software is Primavera. It has different versions, Primavera Project Planner, and Primavera Enterprise.

We have selected a construction project for our research which is a four-story complex, which encompasses a total area of 52,000 square meters, combines the elegance of traditional Iranian architecture with the advanced techniques of modern construction. The Complex is designed with the aim of providing all the conveniences that a modern trading center might need, including state-of-the-art communication links, restaurants, sport, recreational and parking facilities [9].

The PMO office stores information of each project which is gathered by Primavera software in the computer data repositories. These accumulated data contains a vast amount of information for analyze and assess. These data contains everything about projects, like data about activities and resources. In addition data has been saved every week, so we have weekly snapshot of the projects, which means we can analyze trends over time series.

We aimed to use data mining techniques to assess and analyze the data gathered by PMO office in order to find meaningful information; this information could be interpreted to future project management teams and project. We used CRISP-DM[2] methodology for standardization of our data mining project. It has got six steps for a typical data mining project. Also data mining can offer some techniques for predicting the future status of projects, for example giving some old project's data to it, it can predicts the situation which we are interested to them. Delay is one of our main variables to discuss. It is one of the main indicators of the project time management and somehow it shows the quality of time management in projects.

Data which are used for data mining in project management could come from different places, there are some traditional systems in which all data about project time is gathered on the papers in the shape of Gantt chart, on the other hand as we have mentioned some other project management teams have access to project management software packages which can record events more than just project time area. Also some project management software packages are integrated and are working over the network, and are able to record almost every transaction inside the project.

This paper is oriented to specific application of data mining in project management (particularly project time management). It is arranged in the following way: section 2 describes the general process of data mining in project management, section 3 details the CRISP-DM cycle, and its deployment; at the end conclusion and further research has been outlined.

2. Process of Data Mining in Project Management

Project management team, at the beginning of a project - particularly construction projects - starts their work with activities duration estimation, which they may use some experts in this field to estimate activities duration based on the allocated resources to that specific activity. Also they may calculate other resources like cost and labor. They change the time of activities' start and end in a way that the levels of resources become almost as constant as possible. In the real project after the starts of the project, due to some factors (external and internal) some project activities might have some delay in completion, so the activity network may cause other successor activities to start with some delay.

Now due to huge amount of the number of activities and their precedence relations in activity network in different projects, we should use computer to analyze and mine related data. We process the data gathered by project management software and put them in data mining algorithms to see

[2] Cross-Industry Standard for Data Mining

the patterns and relations. Most of the project management software packages store their data in the format which just could readable only by their software. This means that the only way to access their data is to use that specific software. On the other hand most of the data mining software solutions are able to connect to the specific data files, for example Microsoft Project can connect to MS SQL Server to store its data or Primavera Enterprise should connect to MS SQL Server. This requires us to have a data transformation from specific format to another one. We deal with this task later in the preprocessing section.

As we have mentioned, we used CRISP-DM [10] methodology to standardize our work in data mining, CRISP-DM has 6 steps for a cycle of data mining. Data mining process is a cycle and in each cycle these 6 steps should be followed. You can see the dynamics of CRISP-DM methodology in figure 1.

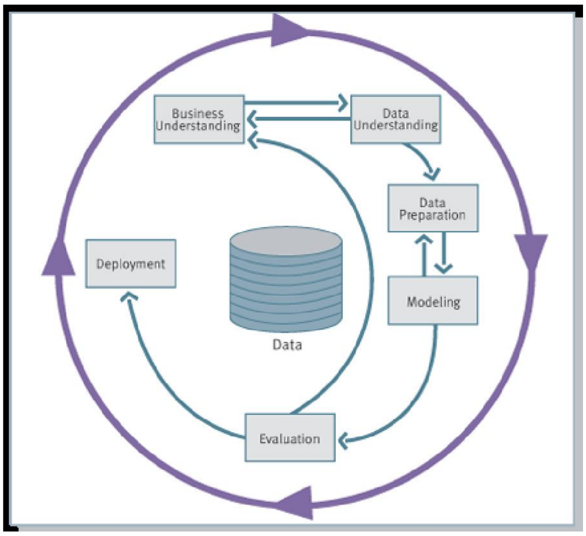

Figure 1: Phases of the CRISP-DM Process Model [11]

3. Data Mining Procedure in Project Management based on CRISP-DM

In this section we discuss the six steps and their goals and objectives:

3.1. Business Understanding: it is also called research understanding phase [12]. This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives [11]. Our goal is to support project managers or project management team by showing them pervious projects data dynamics and then providing better ground for future project's time planning. This requires us to gather some old project management time data, after prepossessing steps by using some data mining models, techniques and tools we would be able to analyze them. When we analyzed data it can be considered as the company process asset and could be used in the future projects. It should be considered that analyzing meaningful information from data mining requires a multi knowledge team.

The project is a huge size construction project (52,000 square meter construction) and therefore a small amount of improvement in productivity would results in saving a lot of money for projects stakeholders. Project management methodology for this project was based on PMBOK, which look at the project from various views, and in this atmosphere we will focus on the time management issue. As a result we would be able to separately analyze time related data easily. Another important point is that project management team has used Primavera as project management software to manage and control the project status including time, human resource, etc.

3.2. Data Understanding: The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information [11]. In this phase we gather needed data in two steps. In first step after negotiation with some members of project management team, they gave us a snapshot of their project which let us know information about activities such as “percent completion”, “activity sequence”, and “activity precedence”. The data has been transformed to the understandable format.
The following is the road map which we have used for data transformation.
Step 1: Transforming Data from Primavera into SQL Server: Primavera Project Planner software uses files with P3 extension to store its data; usually there is a folder that contains several P3 files. Each folder belongs to one project. Actually it consists of one project plus its target. Project management team makes a backup each week just by copying the folder that has projects P3 files. Figure 2 shows a view of data structure that we have got for data mining project.

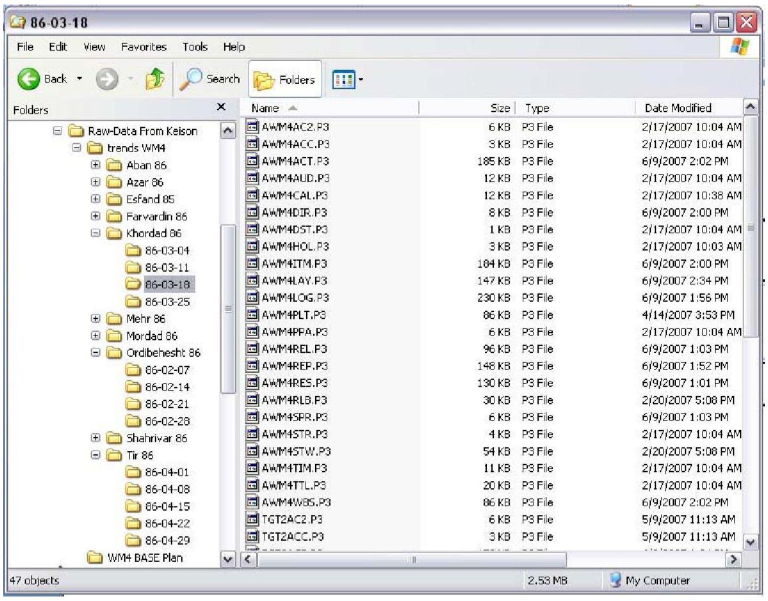

Figure 2: Structure of Raw Data

In the first step we imported Primavera Project Planner’s P3 files into Primavera Enterprise; this automatically saves data in MS SQL Server, because Primavera Enterprise should be connected to MS SQL Server, and it saved every data about project in SQL Server. We should import every projects data into Primavera Enterprise separately. This means that for each snapshot of project we should import data manually. From the beginning of the project a target has been considered for it, which showed in future what should be the status of the project at the end of each week. It is common for many construction projects that after about ten months they revise their target goal, so we would have three separate project’s data in our final weeks of project which belongs to actual status, preliminary target goal and revised target goal of project. In figure 3 you can see the details of this step; AWM4 mentions the actual status of the project, TGT2 denotes to preliminarily project target goal and TAR2 denotes to projects revised target goal.

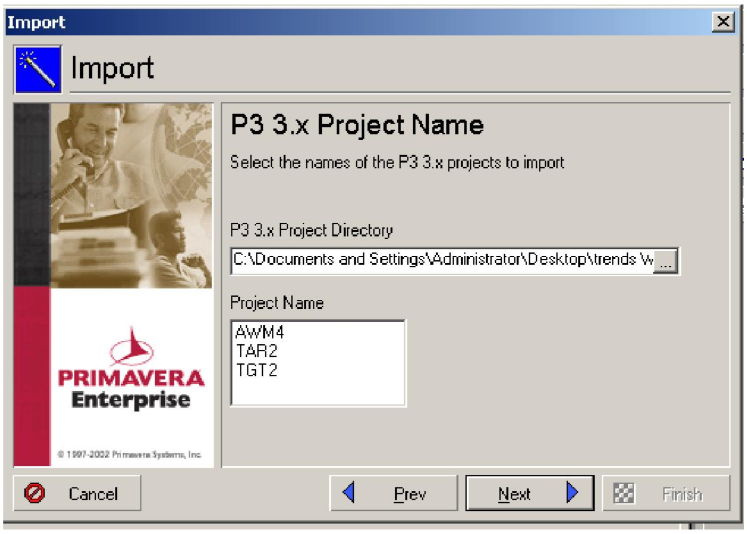

Figure 3: Three projects of a Primavera Project Planner folder

Step 2: Extracting Data from SQL Server Environment into Excel Software: In spite of the fact that most of data mining software packages could connect to MS Excel to input their data, we first transform data from MS SQL Server to MS Excel software, and then we will connect Clementine to these excel files to input their needed data. This is because that Spread Sheet software packages prepare a simple environment so that first, have an insight about the shape of data, and second, having some simple transformation in data. We had chosen MS Excel because it could be simply connected to MS SQL Server and it is one of powerful software in this area. Figure 4 shows a view of MS SQL Server's Enterprise manager software that has our final data set in it.

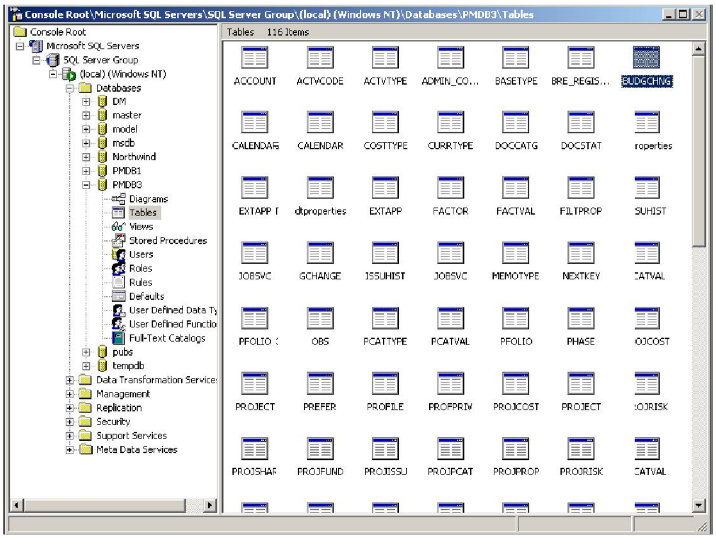

Figure 4 – Enterprise Manager Software

Then by using a Query in Enterprise Manager environment we could save our final data set in Excel format. Here you can see our final query which was run.

```
SELECT     privuser.TASK.task_id, privuser.TASK.proj_id, privuser.TASK.task_code,
privuser.TASK.task_name, privuser.TASK.phys_complete_pct ,
```

```
                      privuser.TASK.status_code, privuser.TASK.total_float_hr_cnt
/ 8 AS Total_float, privuser.TASK.free_float_hr_cnt / 8 AS Free_float ,
                      privuser.TASK.remain_drtn_hr_cnt / 8 AS remain_duration,
privuser.TASK.target_drtn_hr_cnt / 8 AS target_duration,
privuser.TASK.late_start_date ,
                      privuser.TASK.late_end_date, privuser.TASK.early_start_date,
privuser.TASK.early_end_date, privuser.PROJECT.proj_short_name ,
                      privuser.PROJECT.last_recalc_date,
privuser.PROJECT.last_tasksum_date, privuser.ACTVCODE.actv_code_name ,
                      privuser.ACTVCODE.short_name
FROM          privuser.TASK INNER JOIN
                      privuser.PROJECT ON privuser.TASK.proj_id =
privuser.PROJECT.proj_id INNER JOIN
                      privuser.TASKACTV ON privuser.TASK.task_id =
privuser.TASKACTV.task_id AND privuser.PROJECT.proj_id = privuser.TASKACTV.proj_id
INNER JOIN
                      privuser.ACTVCODE ON privuser.TASKACTV.actv_code_id =
privuser.ACTVCODE.actv_code_id
WHERE      (privuser.TASK.proj_id > 389) AND (privuser.TASK.proj_id < 408) OR
           (privuser.TASK.proj_id > 409)
```

Then by using Data Transformation Service (DTS), and having above query we reach to our goal in this step. In figure 5 you can see a snapshot of DTS when we were going to save our data in MS Excel format.

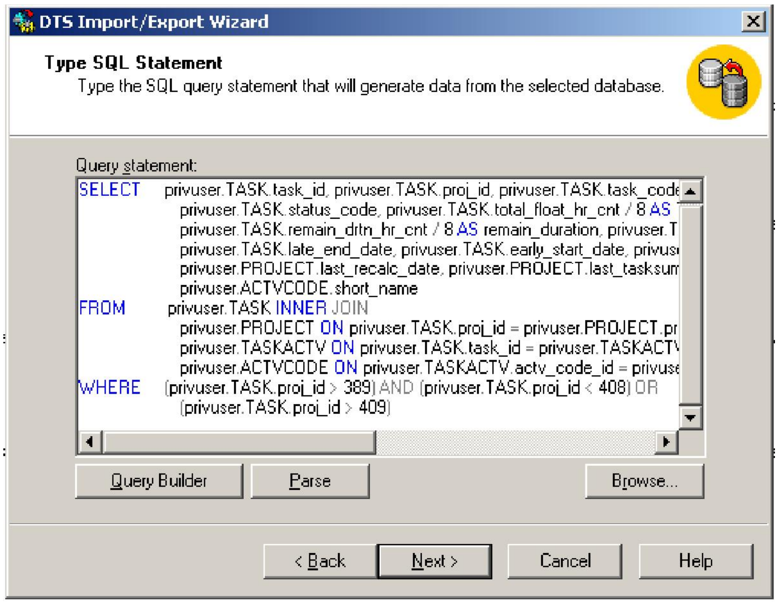

Figure 5 – Data Transformation Service

The data in raw excel file has the structure as mentioned in the tables 1.

Table 1 – Structure of raw excel file

No	Field	Type	Min	Max	Mean	Std. Dev	Unique
1	Week_No	range	1.000	40.000	21.324	11.288	--
2	task_code	discrete	--	--	--	--	--
3	task_name	Set	--	--	--	--	134
4	phys_complete_pct	range	0.000	100.000	62.194	46.300	--
5	status_code	Set	--	--	--	--	3
6	Total_float	range	-107.000	292.000	59.033	72.257	--
7	Free_float	range	0.000	265.000	20.521	44.601	--
8	remain_duration	range	0.000	283.000	17.777	39.587	--
9	target_duration	range	0.000	120.000	13.100	10.634	--
10	late_start_date	range	2006-09-16	2008-01-09	--	--	--
11	late_end_date	range	2006-09-17	2008-01-08	--	--	--
12	early_start_date	range	2006-09-16	2008-03-21	--	--	--
13	early_end_date	range	2006-09-17	2008-03-25	--	--	--
14	proj_short_name	Set	--	--	--	--	37
15	last_recalc_date	range	2006-09-16	2007-12-07	--	--	--
16	Flor	Set	--	--	--	--	8
17	Function	Set	--	--	--	--	6

As you can see in table 1, all the actual project status and its target and revised target in every week in the project life cycle is in the same columns, so we need more transformation to keep each activities in a one record. We cannot use MS Excel for this purpose. In Clementine we have more tools to transform data structure.

Step 3: Final Data Preparation in Clementine Software: as we have mentioned above, first we should gather each activity's data in a one record, so we used Merge tool in Clementine Environment. After that we continued to modeling our data mining problem in Clementine software. By using Type tool we would be able to mention the type of data in each column. Also Clementine has got some tools that could even recognize the type of columns automatically but because of quality of these tools we preferred to mention the type manually. In the figure 5 you can see the details of this process.

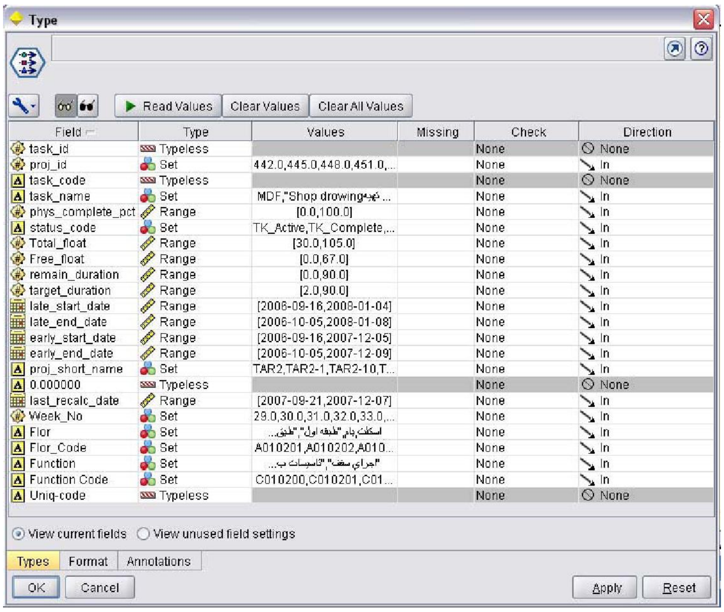

Figure 5 – Details of determining the type of data

Then we would proceed with adding some new fields to the main data, by using some operation based on the other column's data. For example comparing the actual situation of an activity with its target we could calculate the delay that the activity would have.
In this step in CRISP-DM methodology we would like to have a look at the nature of data in order to see the obvious patterns and having an insight to the data, so we will continue our work with Expletory Data Analysis. Basically when a data miner has a strong insight into data he could continue other steps with more confidence and better quality. As it is obvious from its name at this time we are not looking for something specific, but just exploring data to see its structure and clear trends and patterns. Clementine has got so many easy tools just for Exploratory Data Analysis, such as data audit [13]. We have used these tools and their results also were very helpful, even some of results were shocking. For example the delay distribution versus activity total float has got two distinct distributions, one of them was in negative side and the other one was in positive side, in figure 6 you can see the diagram. Also another fact is that the delay is most in the middle of the diagram. (explain)

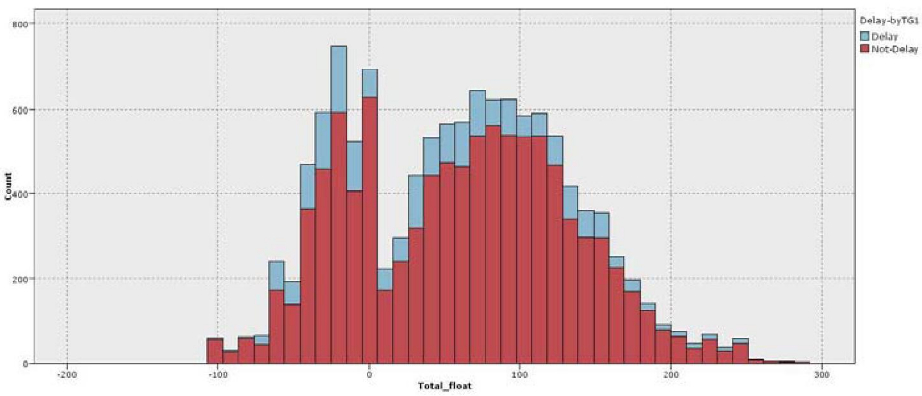

Figure 6 – Diagram of delay versus activity total float

Or, number of delays in first and second floor was more than other floors. Which somehow against our initial expectations, that when a project is going to end the number of delays should be more. In table 2 you could see the detail of this fact in our data mining problem.

Table 2 – Distribution of delay

No	Value	Proportion	%	Count
1	1st Floor		17.06	2,050
2	2nd Floor		17.06	2,050
3	3rd Floor		17.68	2,124
4	4th Floor		17.06	2,050
5	General		6.99	840
6	GroundFloor		17.68	2,124
7	Roof		3.39	407
8	Structure		3.08	370

Delay-byTG1
Delay Not-Delay

Also at the end of this Phase we should have some idea about the quality of data. As we could see the results in the Exploratory Data Analysis we can come to this conclusion that the quality of data is high enough to continue our data mining problem. It was predictable because the type of data that we had from SQL Server would assure us that the quality should be good. In these types of data there is a very low probability of having any mankind mistake.

3.3. Data Preprocessing: The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modeling tools [11]. We transform raw data in a format that it can be readable by data mining software. This could be one of our major achievements in this paper and it will be explained later in this paper.

The main process was described in 3.2, but in this phase we add some additional steps to construct our final data set. Moreover some data mining algorithms need specific type of data as their input. For example X Algorithm needs discrete data as input, continues data fails to input to this algorithm. This step and step 3.2, could be very time consuming, and it has been mentioned that these steps could take more than 60% of data mining project time [14]. On the other hand because our data quality is high, the time for data preparation decreases.

Most of our work in this step is adding some additional fields to main data, for this purpose we have used required nodes in Clementine software. In figure 10 you may see one of main node set in Clementine software during the modeling of our data mining problem.

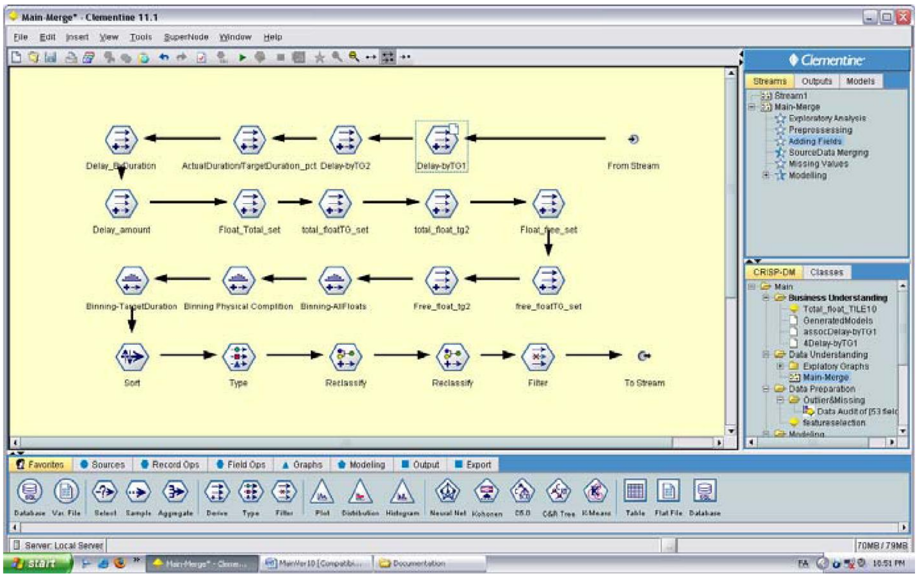

Figure 10 – Clementine Software Node Design

We calculate delay with two bases, one of them is first planned target goal and second is revised target goal, so we add these two fields to our main data set. We also used Binning[15] tool in Clementine for changing data type for some fields from continues to discrete, like duration which has got more than 300 values, we divide it into 10 sets. Most of data mining algorithms requires data miner to determine which fields are input and which one is output, we did it by using type [16] node in Clementine, figure 11 shows a typical view of type node in our data mining project. As we have discussed since our major goal in time management knowledge area is to analyze delay, we have marked delay as our output variable.

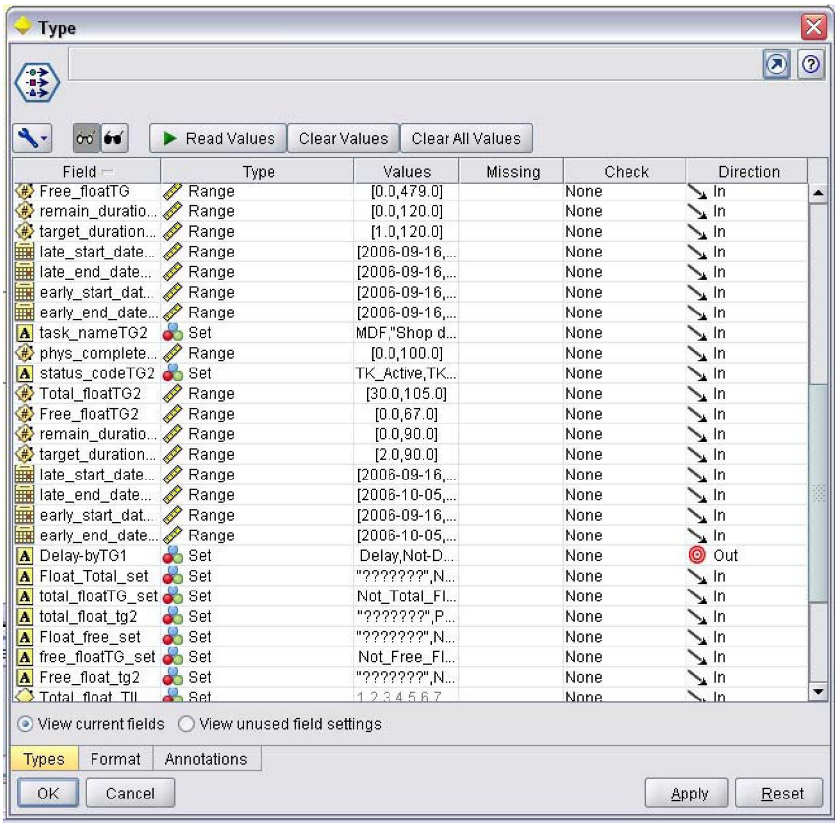

Figure 11 – Type node in Clementine

Another task that we should deal with in this step is *Missing Data*, because our main source for getting data was a computer data base, so the quality of data is very high and we had no problem in this phase. Data mining professionals had suggested more than ten ways to fill blank data in our main data set [17]. There is just one case which we would see the missing data in our data, as we have mentioned before, after about 10^{th} month of projects start, project management team gives revised target plan and we have entered its data in separated fields in our main data set, but for activities which has been done in advance of 10^{th} month, those fields are blank and Clementine filed those fields with *$null$*. It is clear that the nature of these missing data is quite normal and it has got nothing to do with process of our data mining problem. But we should take care of these values in next step by designing our models.

Another major task that we should take a care of it in this step is data transformation, in MS SQL Server durations has been saved in hour unit, because in most of project management teams in Iran, specialists works with day unit, so we would need a transformation for duration and float fields. Actually we have done this step in our final query which has been mentioned above.

3.4. Applying Data Mining Models: In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed [11]. In Phase III we have some transformation of data in a way that it can be fed into specific type of data mining techniques. Then in this phase we used some data mining techniques which seem to be more appropriate for our data and our goals.

Data mining has got some major tasks, like data visualization, classification, clustering, association rules etc. Due to the structure of our data, and the goals that we are looking for it, we have used classification and association rules techniques in this data mining problem. Clementine software also supports these techniques very strongly.

3.4.1. Classification: Classification is a mapping from a (discrete or continuous) feature space X to a discrete set of label Y [18]. Classification predicts class labels. This is a supervised classification which provides a collection of labeled patters, the problem being to label newly encountered, still unlabeled, patterns [4]. In project management data mining, classification may use for determining the most similar activities in a specific groups, which allows project management team to study those activities together for managing delays better, so in classification models we label delay as our output variable which means that we want to classify all records regarding their delay field.

Clementine has a lot of classification algorithms like C5 and CHAID. C5 has used in our data mining problem for the mentioned purpose; here you can see a brief result of this algorithm.

```
if Float_Total_set = Negative_Total_Float and Floor = 1st Floor and status_code = TK_NotStart then Delay
if Float_Total_set = Negative_Total_Float and status_code = TK_Active then Delay
```

Some of these rules may seems to be very obvious, in data mining these kind of results are normal and data miners should take care of them. On the other hand these obvious rules show that the logic inside the data is correct, and verifies the quality of our data.

CHAID classification is another algorithm which has used beside C5. In the figure 12 you can see the tree structure of a result of CHAID classification algorithm.

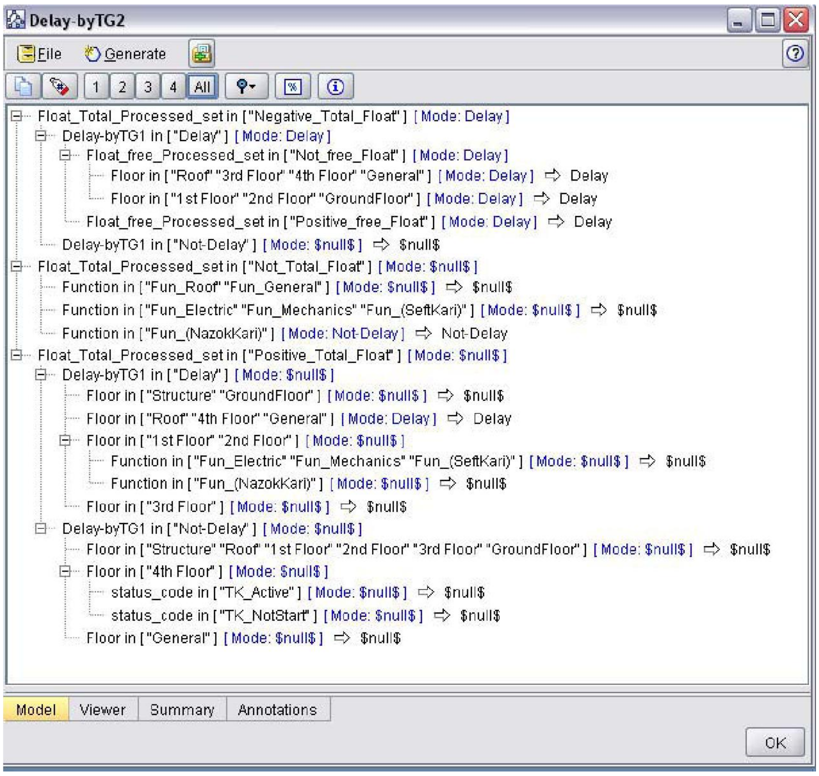

Figure 12 – CHAID Output

Beside of C5 and CHAID we have used QUEST and Decision Tree tools to classify our data in way that we become closer to our main goal which is defined in Phase I.

3.4.2. Association Rule Mining: Association rule mining discovers relationships among attributes in data bases, producing IF THEN statements concerning attribute-value [4]. An X=>Y association rule expresses a close correlation between items in a database with values of support and confidence. The confidence of the rule is the percentage of transactions that contains the consequence in transactions that contain the antecedent. The support of the rule is the percentage of transactions that contains both antecedent and consequence in all transactions in the database In Table 3 and 4 you can see a summary of association rule mining; there are also a lot of uninteresting rules, like the great number of redundant rules (rules with a generalization of relationships of several rules). There are some similar rules (rules with the same element in antecedent and consequent but interchanged). And there are some random relationships (rules with random relations between variables). But there are also rules that show relevant information for educational purposes. We have used Apriopri and GRI as our association rule mining algorithms, in table 3 and 4 you can see their details

Table 3 – Apriopri Results

No.	Consequent	Antecedent	Support %	Confidence %
1	Delay-byTG1 = Not-Delay	Function = Fun_Electric and Float_free_set = Not_free_Float	21.29	99.062
2	Delay-byTG1 = Not-Delay	Function = Fun_Electric and Float_Total_set = Positive_Total_Float and Float_free_set = Not_free_Float	15.63	98.722
3	Delay-byTG1 = Not-Delay	Function = Fun_Mechanics and Float_free_set = Not_free_Float	16.787	95.34
4	Delay-byTG1 = Not-Delay	Function = Fun_Mechanics and Float_Total_set = Positive_Total_Float and Float_free_set = Not_free_Float	12.659	94.938
5	Delay-byTG1 = Not-Delay	Function = Fun_Electric and Float_Total_set = Positive_Total_Float	23.928	92.209
6	Delay-byTG1 = Not-Delay	Function = Fun_Electric	30.445	91.334
7	Delay-byTG1 = Not-Delay	Function = Fun_Mechanics and Float_Total_set = Positive_Total_Float	18.494	90.639
8	Delay-byTG1 = Not-Delay	Function = Fun_Mechanics	22.946	90.606
9	Delay-byTG1 = Not-Delay	Float_Total_set = Positive_Total_Float and Float_free_set = Not_free_Float	50.861	87.94
10	Delay-byTG1 = Not-Delay	Float_free_set = Not_free_Float	76.33	87.231

Table 4 - GRI[3] Results

No	Consequent	Antecedent	Support %	Confidence %
1	Delay-byTG1 = Not-Delay	Delay_amount < 1.000	83.32	100.0
2	Delay-byTG1 = Not-Delay	Delay_ByDuration	78.98	100.0
3	Delay-byTG1 = Not-Delay	Delay-byTG2 = $null$ and Delay_ByDuration	54.85	100.0
4	Delay-byTG1 = Not-Delay	ActualDuration/TargetDuration_pct > 6.667	51.57	100.0
5	Delay-byTG1 = Not-Delay	remain_duration < 1.000	51.49	100.0
6	Delay-byTG1 = Not-Delay	status_code = TK_Complete	51.48	100.0
7	Delay-byTG1 = Not-Delay	phys_complete_pct > 99.500	51.48	100.0
8	Delay-byTG1 = Delay	status_codeTG = TK_Complete and ActualDuration/TargetDuration_pct < 6.667	16.7	99.9
9	Delay-byTG1 = Delay	Delay_amount > 1.000	16.68	100.0
10	Delay-byTG1 = Delay	status_codeTG = TK_Complete and Delay_amount > 1.000	16.68	100.0

3.5. Evaluate Data Mining Models: At this stage in the data mining project you have built models that appear to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached [11]. Information and results of data mining techniques has been observed carefully by the project management team, some trends have got logical reasons. On the other hand by using some evaluation tools in data mining software we can see that quality of results are high enough.

Clementine has got powerful tools for evaluating models that have been used in previous phase, for example the evaluation result of our first classification model is as … table ….

Table X – Evaluation of First Classification Model

```
Results for output field Delay-byTG1
Comparing $C-Delay-byTG1 with Delay-byTG1
Correct              11,187   93.11%
Wrong                   828    6.89%
Total                12,015
```

[3] Generalized Rule Induction

From the table X you can see that in 93.11% of times the classification model has done its job correct, which is high enough to make this classification model reliable. There were other classification and association rules mining models, also their evaluation showed high amount of confidences with in data. When the confidence level of generated models are high, it shows that there would be patterns in data, otherwise when there are no rules generated or the confidence level of generated rules are low so there is not any strong pattern or trend inside data. In table X and Y you can see the confidence percentage and support percentage for two association models that have been showed in previous section.

3.6. Implement Data Mining Model: Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases it will be the customer, not the data analyst, who will carry out the deployment steps. However, even if the analyst will not carry out the deployment effort it is important for the customer to understand up front what actions will need to be carried out in order to actually make use of the created models [11]. In our data mining project we generate a report for Kayson Company and we taught them how to use them. Also we introduced the dynamics of CRISP-DM cycle to them and we let them know they can introduce other new data mining projects based on the results of this project.

4. Conclusion

This is one of first data mining projects which are done in project management area, and in very first results it revealed that there are some strong patterns and relation between data. These relations belong to the nature of projects, and their discovery may have important effects in the development of project management knowledge. For simplicity we organize our conclusion in these categories:
First, the results of above study has showed that there are some strong patterns in the data that should be study by the appropriate specialists, it could be continue even with defining new data mining projects as a CRISP-DM cycle. It is recommended that data miners input other variables which are belong to other domains of project management, like cost, human resource, etc. Second, more especially in the results of this data mining project we achieve to the some interesting and important patters and trend in time management area, some of them has been showed in 3.4 section of this paper. Third, we have introduced a simple and efficient way of data transformation of data, suitable for data miners who are working in project management area. It was one of time consuming tasks in this data mining project. Forth, prediction is a tool in the hand of management, and data mining could be used as one of strong tools for prediction. So we could use data mining as a prediction tool for project managers who want to predict the status of their projects based on the data of old projects.

5. Future Work

Data mining is an analytic science and when it has done generally its results could have dependable. In addition as we have mentioned previously in CRISP-DM methodology every cycle of a data mining project could result in other date mining projects[10], so here we will describe some of suggested projects:
5.1. General Construction Project Data Mining: In spite of the fact that time management is one of core domains of project management [6], have a data mining project in just one domain is not general, so it should be done beside other domains, like cost and human resource. The relation

between time and HR[4] in project management is strong. When ever in a project an activity has got delay and project manager has the idea to speed up that activity, he may inject more resources like labor to decrease duration of the activity. So the relation between time and HR indexes and factor should be study closely.

5.2. Forecast of New Projects: It could be useful to have some studies in order to see the results of forecast task of data mining.

5.3. Relation of different Construction Projects: It could be useful to see the results of data mining projects in different areas of construction. Building a middle size marketing place or a Dam on the river even constructing a single little house may have a lot of differences in patterns and trends in data.

6. References

1. Garther Group, www.gartner.com
2. Rachel Konrad, Data Mining: Digging user info for gold, ZDNET News, 2001, Feb the 7, 2001
3. The Technology Review Ten, MIT Technology Review, January/February 2001
4. Cristobal Romero - Sebastian Ventura - Enrique Garcia, Data Mining in course management systems: Moodle case study and tutorial, Computers & Education, 51, 2008,
5. Danial T. Larose, Discovering knowledge in data An introduction to data mining, Hoboken New Jersey, A john Wiley& Sons, Inc publication, 2006,
6. Project Management Institute, Project Management Body on Knowledge – PMBOK, 3rd edition, Pennsylvania, Project management institute, 2004
7. Peter Morris, The Management of Projects, Thomas Telford
8. Rory Bruke, Project management – planning and control techniques, forth edition, Great Britain, John Wiley, 2004, Page 7
9. Kayson Company, www.kayson-co.ir
10. www.crisp-dm.org
11. http://www.crisp-dm.org/Process/index.htm
12. Danial T. Larose, Discovering knowledge in data An introduction to data mining, Hoboken New Jersey, A john Wiley& Sons, Inc publication, 2006, Page 6
13. Clementine 11.1 Application Guide, 2007, Integral Solution Limited, Page 23
14. Daniel T. Larose, Discovering Knowledge in Data (An Introduction to Data Mining), 2004, Wiley, Page 28
15. Clementine 11.1 Application Guide, 2007, Integral Solution Limited, Page 41
16. Clementine 11.1 Application Guide, 2007, Integral Solution Limited, Page 101
17. Daniel T. Larose, Discovering Knowledge in Data (An Introduction to Data Mining), 2004, Wiley, Page 28
18. Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Pattern classification. Wiley Interscience.

[4] Human Resource