We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 3

Determination and Classification of Crew Productivity with Data Mining Methods

Abdullah Emre Keleş and Mümine Kaya Keleş

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75504

Abstract

Turkey is a developing country and the main axis of development is "construction." The construction sector is in a position to create demand for goods and services produced by more than 200 subsectors, and this widespread impact is the most basic indicator of the sector's "locomotive of the economy." In the development of the construction industry, crew productivity plays a very important role. While businesses that do not measure their employees' needs, their locations, and so on are suffering from various losses, rare businesses that take these parameters into account can profit. The identification of leadership types that will motivate employees has great importance in terms of construction businesses where the human element is the foreground. For this purpose, in the province of Adana, the relationship of productivity between the engineers working in construction companies and workers who work at lower departments of these engineers was examined. In this study, bidirectional multiple leadership questionnaire (MLQ) was applied to construction site managers and employees, and according to this survey data, leadership and motivations/productivities were classified using data mining methods. According to the classification analysis results, the most successful data mining algorithm was random forest algorithm with a rate of 81.3725%.

Keywords: classification, construction information, construction management, crew productivity, data mining, random forest algorithm, supervised and unsupervised learning, Weka

1. Introduction

With the increasing globalization in the construction sector, institutionalization is at the forefront. In addition, under increasing competition conditions, construction companies

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

are forced to differentiate in the methods and technologies they use in business processes in order to be able to share in the sectoral market and to protect and strengthen it [1]. When the construction works are examined in terms of their management functions, the result is that they are still largely human-focused. This situation pushes construction companies to survive in a highly competitive environment and to take positions in construction businesses, especially by increasing their loyalty to operating by transferring employees to business processes with greater motivation, and thus to obtain more efficiency. In short, the human factor, one of the inputs used in construction production, must be managed correctly. Businesses have to deal with systematic approaches through the management of employees.

When the subject of this study and similar studies are examined, it can be seen that the studies related to construction project management give different and highly motivated results in the regions where the study is conducted, but they can give different results when the same study is conducted in another region. These researches have been inspired to make this study. The situation in Adana province, the relationships between the leadership styles and the motivations of employees are determined and only the results based on this province are taken into consideration in terms of the enterprises and employees working on this province.

The determination of the productivity of workers in the construction sector is directly related to the success of the enterprises. In the face of increasing competition, the businesses that do not measure the needs, locations, activities, and so on of their employees are also undergoing various losses, even if they are unaware.

In the light of all these, the main objective of this study is to determine the relationship between the civil engineers who are in managerial leadership position in the construction enterprises operating in Adana province and the subordinates of the master-worker-headman positions they work with in the leadership-motivation axis. In line with this aim, 100 construction companies were selected to conduct construction projects in Adana province and two questionnaires were applied on two sides, one for the leader and one for the employee.

For this purpose, the productivity relationship between the persons who are engineers work in the city of Adana, the ones who produce the building, the ones who work in the construction companies and workers who work in the subhierarchy of these engineers were examined. The identification of leadership types that will motivate and support employees has a great importance in terms of construction businesses where the human element is the fore ground. From the point of view of the construction site managers in charge of the sites, it is thought that it will be useful for the sector representatives, businesses and all employees to determine which leader type will motivate which employees. In this context, association rule mining were made with Apriori Algorithm using data mining methods using Weka [2] and Keel [3] software on the data obtained by the multiple leadership questionnaire (MLQ) applied bidirectionally to site chiefs and employees, and leadership and motivations/efficiencies were classified by using classification algorithms. The impact of leadership styles

on employee motivation/productivity has been analyzed. Thus, it is aimed to present the creation of the most suitable rules that can be used in the field of engineer leadership-employee motivation/productivity of the construction companies in Adana province and to present them to the sector.

The results obtained by this method are analyzed with these data mining methods, and they are given and interpreted in "Findings and Discussion" and "Conclusion" sections of this study.

2. Literature review

When the literature on the topic is considered, it is determined that there are deficiencies in the "Systematic Leadership Approach in the Construction Sector". It has also been determined that there is not enough consciousness about what employee productivity will be when applying which leadership styles. At the same time, it has been determined that there are deficiencies in how to increase employee productivity on a sectoral basis. Even if the number of works based on the construction sector in the field of employee productivity has increased in recent years, there is little research to measure the productivity by evaluating both sides and the subordinate relationship. It is believed that it is important to provide a systematic productivity analysis to the workers in the construction sector by eliminating this deficiency. Such a systematic development and submission to the use of sector representatives may lead to the deficiencies mentioned earlier and may also provide guidance.

Unfortunately, too many sources were not found when the literature review was conducted in this area. Kaya and his colleagues [4] tried to estimate the productivity values with the help of data obtained from the survey of ceramic workers working in construction companies. These values, which are also used as attributes in the related study, are the number of teams, the work experience and the average age of the people on the crew. In their study, it was focused on how to achieve high productivity in ceramic works by means of mining rules by classifying productivity with the ceramic data obtained by using measurement methods. Andaç and Oral [5] presented the results obtained from their work using the Artificial Bee Colony Algorithm for estimating worker/labor productivity. In the work of Keleş and Kaya [6], they used the association rules and a data mining method, in order to increase employee productivity with the demographic information obtained from the workers working as wall masters/mason in the construction sector. In Keles's PhD thesis [7] "Determining the relationship between leadership of the site managers and motivation of their employees with the data mining in construction projects," a double-sided survey has been applied to site managers and their subordinates who work in the construction sites in Turkey. Following this conducted questionnaire, leadership was identified from two different perspectives. After determining the leadership models, the relationship between the leadership of the site chiefs and the motivations of the workers was determined by using the association rule mining method from the data mining methods.

3. Materials and methods

In this chapter, the changes of the productivity of the other employees are discussed according to the leadership models of the construction engineers working as construction engineers. In this respect, data will be obtained from the point of view of the site managers and employees. In other words, it will be determined that, productivity will increase as a result of which modern leadership types—transactional/transformational/passive avoidant leadership behaviors—in the literature studies especially in recent years, are applied to the construction worker/employee group in which characteristics.

In this study, as a method of obtaining data, questionnaires were used to provide "bidirectional evaluation" in order to reach the targeted results realistically. In other words, not only the behaviors, characteristics and the like of the leaders, but also the factors such as expectations, characteristics and style of living are taken into account.

After this data which will be obtained from the construction companies in Adana province through the questionnaire forms, data mining studies have been started. Detailed survey studies for the determination of relations have been applied to civil engineers working in construction companies as construction site supervisors and other employees in construction sites. A bidirectional model designed to be tested in this context is shown in **Figure 1**.

When the relevant model is considered, it has been decided that the implementation of a bidirectional questionnaire will be positive and appropriate, as explained earlier. When all theories and methods in the relevant literature are examined, it was found appropriate to use the multiple leadership questionnaire (MLQ) scale developed by Bass and Avolio [8] in this study to determine the types and characteristics of the engineers, as shown in **Figure 2**.

Information was gathered from the site chiefs about the efficiency value to be calculated for employees who will evaluate themselves. These collected data were added to the end of MLQ surveys applied to engineers. The chiefs assessed the productivity level of their employees by choosing between low, medium and high. In the employee questionnaires, employees indicated their productivity information by choosing one of the low, medium and high options according to the management of the site chiefs and leadership understanding. It is ensured that the data mining methods that form the basis of the study in this way are consistent in the data to be applied.

Since MLQ has a bidirectional survey application system, it reveals how leaders perceive the way of management both from the point of view of themselves and from the point of view of employees. Forty five questions in the short measure are asked to the leaders in active mode and to the employees in passive mode. The relevant scale with this feature, that is, the same type of questions are directed to the people on both sides of the subject, it is possible to obtain more healthy results. With this feature, MLQ is a leadership survey questionnaire that has been used in recent years by many researchers in different disciplines. The survey questionnaire used in this study does not include personal information of the persons, only

http://dx.doi.org/10.5772/intechopen.75504

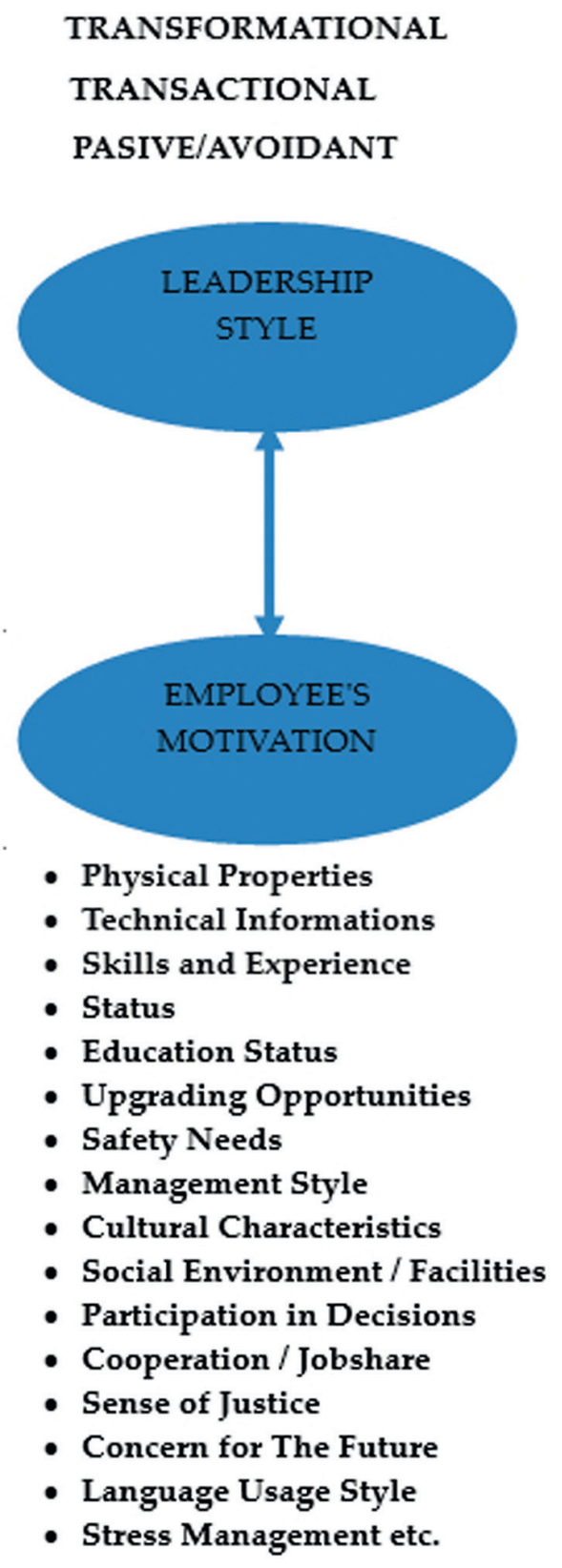


Figure 1. Model designed to be tested.

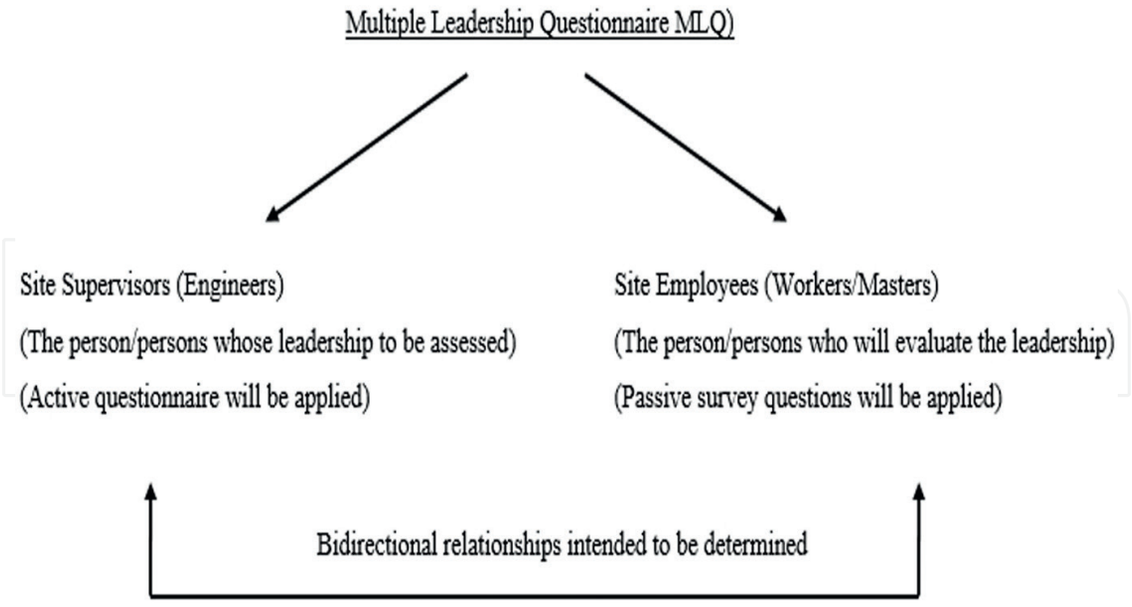


Figure 2. Systematization of the related survey study and the goals planned to be achieved.

information like their age range, their gender, how much they are working, and so on were collected. In this study, a scale covering 45 questions, which was revised and abbreviated, was used instead of 72 questions.

According to the main axis of the study, data mining methods are used together with Weka and KEEL programs for classifications. In a sense, this study is supported by a different perspective that is not frequently used in the sector, and the sector has benefited.

Data mining is the process of extracting previously undiscovered information based on a wide variety and quantity of data held in data warehouses and using them to make decisions and action plans. It is the search for relationships and rules that will allow us to make predictions about the future from a large amount of data. Data mining is the semiautomatic discovery of patterns, relations, changes, irregularities, rules and statistically significant structures in data. The computer is responsible for determining the relationships, rules and properties between the data. The goal is to be able to detect previously unrecognized data patterns [9].

It is necessary to address the different types of data for an effective data mining application, to ensure the effectiveness and scalability of the data mining algorithm, to provide usefulness, accuracy and significance of the results, to display the discovered rules in various forms, to process data in different environments and to provide privacy and data security features. Alternatively, data mining is actually regarded as a part of the knowledge discovery process. The stages of the knowledge discovery process are as follows [9]:

1. Data cleaning (remove loud and inconsistent data)
2. Data integration (combining multiple data sources)
3. Data selection (determine the data related to the analysis to be performed)

4. Data transformation (to transform the data to be used by the data mining technique)
5. Data mining (implementing intelligent methods to capture data patterns)
6. Pattern evaluation (to identify interesting patterns representing information obtained according to some measurements)
7. Information presentation (performing the user presentation of the obtained information that has been mined) [10, 11].

The difficulty in obtaining the relevant data from the large data generated by the use of technology in every sector is also valid for leadership. In order to obtain meaningful and useful information from meaningless data heaps, it is planned to use data mining methods in this study. For this reason, the data in the surveys gathered within the scope of the study were primarily preprocessed and then prepared in the relevant file format for operation in data mining programs. Today, both commercial and open source programs have been developed to make data mining studies. There are many algorithms in these programs. By using these algorithms, meaningful information can be extracted from the data available [9]. In this study, KEEL software was used for preprocessing steps and Weka software for classification steps.

Weka [10] is the abbreviation for Waikato Environment for Knowledge Analysis. It is a Java-based data mining and machine learning software developed under the GNU general public license at Waikato University in New Zealand. It includes preprocessing, classification, clustering, association rule mining, feature selection and visualization processes on data sets. Weka works with the Attribute Relationship File Format (.arff) file format. This file format is a specially designed file format that is kept in a text structure. The @relation, @attribute and @data statements are used to specify the file structure. The @relation specifies the general purpose or name of the stack data. While @attribute is used to specify attribute names that correspond to columns in the data set, @data marks the beginning of the raw data set.

KEEL is software written in Java language developed by the University of Granada with the support of the National Science Projects Agency of Spain. KEEL is not rich in terms of classical data mining algorithms such as clustering. Instead, Fuzzy classifiers, artificial intelligence-based classification and rule-based clustering algorithms are included [12]. One of the weakest software in terms of data visualization is the KEEL. Since KEEL software provides highly advanced algorithms in preprocessing parts according to other software, KEEL software was used in the preprocessing phase of the data obtained from the questionnaires in this study; that is, the data preprocessing level, which constitutes the first four steps of the information discovery process, was realized with the help of KEEL software. During this preprocessing phase, normalization is performed, and the data are transformed into the related form.

4. Findings and results

In the scope of the study, 102 employee questionnaires and 102 leader questionnaires were applied. The results of the obtained leadership outputs are given in **Table 1**.

	Transformational leadership	Transactional leadership	Passive avoidant leadership	Total number of surveys
According to the employee	81	20	1	102
According to the leader	84	18	0	102

Table 1. Survey results.

The collected surveys in the scope of this study were primarily brought together in an Excel file. The data stored in Excel format are then converted to the .arff file format, which is a file format of Weka, so that one of the data mining program, Weka, can be run. For this, various preprocessing was carried out with the help of KEEL program and transformation of file format was performed.

In the scope of the study, min-max normalization method, which is widely used, is used. In the min-max normalization method, the largest and smallest values in a group are handled. All other data are normalized to these values. The purpose here is to normalize the smallest value to be 0 and the largest value to be 1 and to spread all the other data to this 0–1 range. The prescribed formulation is shown in Eq. (1). According to this formulation, v is the value

```
@relation bap_calisan_2017
@attribute gorev
{boyaci,demirci,dogalgaz_doseme,formen,isci,kalipci,parke_ustasi,seramikci,sivaci,tesisatci,usta}
@attribute cinsiyet {1,2}
@attribute yas {1,2,3,4,5}
@attribute egitim {1,2,3,4,5,6}
@attribute maas {1,2,3,4}
@attribute tecrube {1,2,3,4}
@attribute benzer_islerdeki_tecrube {1,2,3,4}
@attribute …….
…….
@attribute motivasyon {yuksek,dusuk}
@attribute lideregore_liderliktipi {transformational,transactional,passive-avoidant}
@attribute calisanagore_liderlik_tipi {transformational,transactional,passive_avoidant}
@data
tesisatci,osmaniye,1,2,3,2,3,3,2,3,2,1,3,0,0,1,3,1,2,7,10,1,yuksek,transformational,transformational
sivaci,kahramanmaras,1,3,3,2,4,4,3,3,3,1,3,0,1,0,3,1,2,7,10,2,yuksek,transformational,transformational
sivaci,adana,1,2,3,1,3,4,3,3,1,1,3,0,1,0,3,1,2,7,10,2,dusuk,transformational,transformational1,3,2,4,6,2,1,0,1,0,3,dusuk,D
isci,adana,1,2,3,2,4,4,3,3,3,1,1,0,1,0,3,1,2,7,8,2,dusuk,transformational,passive_avoidant
formen,adana,1,2,3,3,3,3,2,3,2,3,4,1,0,0,3,2,2,6,8,2,yuksek,transformational,transformational
isci,adana,1,3,2,1,4,4,4,3,4,1,3,0,1,0,3,1,2,7,10,2,yuksek,transactional,transactional
kalipci,erzurum,1,3,2,1,4,4,1,3,1,2,1,0,1,0,3,2,3,7,8,1,dusuk,transformational,transactional
parke_ustasi,adana,1,3,2,1,4,4,3,3,2,1,3,0,1,0,3,2,2,6,10,2,dusuk,transactional,transformational
……..
```

Table 2. Categorical .arff file.

to be normalized, whereas v′ is the value to be obtained as the result of normalization. In the formulation, new_min is taken as 0 and new_max is taken as 1.

$$v' = \frac{v - min_A}{max_A - min_A}(yeni_max_A - yeni_min_A) + yeni_min_A \quad (1)$$

The format of the .arff file is different for each data mining method. Some of the data mining methods in Weka only work with numerical data, while others work with categorical data. For example, while there is a greater need for numerical data for classification and clustering algorithms, categorical or nominal data are needed for the algorithm of association rules [7]. Numerical data were categorized by the uniform frequency method. In the equal frequency method, the property ranges are divided into N ranges, and an equal number of pieces of data are held in each range. This method is used because it can work with distorted data. **Table 2** shows the summary of the .arff file format prepared in the scope of the study.

Correctly Classified Instances			83		81.3725 %		
Incorrectly Classified Instances			19		18.6275 %		
Kappa statistic			0.1878				
Mean absolute error			0.2096				
Root mean squared error			0.3433				
Relative absolute error			92.0293 %				
Root relative squared error			103.2619 %				
Total Number of Instances			102				
=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0.857	0.816	0.988	0.894	0.539	transformational
	0.15	0.012	0.75	0.15	0.25	0.528	transactional
	0	0	0	0	0	0.178	passive_avoidant
Weighted Avg.	0.814	0.683	0.795	0.814	0.759	0.533	
=== Confusion Matrix ===							
a	b	c		<-- classified as			
80	1	0	|	a = transformational			
17	3	0	|	b = transactional			
1	0	0	|	c = passive_avoidant			

Table 3. The classification results of random forest.

Within the scope of this study, the leadership qualities that employees qualify for their leaders are classified. According to the analysis results obtained, the most successful data mining algorithm in terms of classification has been random forest algorithm with a rate of 81.3725%. This algorithm is a community learning algorithm. This algorithm generates more than one decision tree during the classification process and thus aims to increase the classification value. Individually constructed decision trees come together to form the decision forest. The classification results obtained are given in **Table 3**.

5. Conclusion

In this study, some results were obtained by applying "Classification" methods of data mining methods. With the use of classification algorithms in leadership, an analysis of the effect of leaders' leadership styles and behaviors on the motivation of site employees has been made in the construction industry.

According to the MLQ licensed questionnaire score, the leadership style tendencies of construction site supervisors were categorized and evaluated in three different ways as "transformational," "transactional" and "passive/avoidant" within the scope of responses obtained from the survey questionnaires in this study.

The following suggestions can be made to the construction sector, construction companies and construction site managers who are the managers of the construction sites in order to overcome the deficiencies found in the subject in the data presented in this study.

Among the important results of the study, it can be said that sector representatives will benefit from the fact that the motivation of employees can be increased if each employee group has different qualifications and if these leaders are selected and implemented in accordance with these qualities.

At the same time, based on the information obtained as a result of this study, it is considered that important shortcomings can be achieved by sharing the results of the construction sector, which can raise the motivation of the construction site employees with the representatives of the construction sector. In this context, it is thought that it would be beneficial to share the results obtained with the site chiefs and employees in the construction companies with the meetings, seminars and similar sessions.

Acknowledgements

This study was supported by Adana Science and Technology University Scientific Research Project Commission (Project Number: 17103018 and Project Number: MÜHDBF.BM.2015-2111), Turkey.

Author details

Abdullah Emre Keleş[1] and Mümine Kaya Keleş[2]*

*Address all correspondence to: mkaya@adanabtu.edu.tr

1 Faculty of Engineering, Department of Civil Engineering, Adana Science and Technology University, Adana, Turkey

2 Faculty of Engineering, Department of Computer Engineering, Adana Science and Technology University, Adana, Turkey

References

[1] Balaban B. The Effect of the Culture on the Motivation of the Workers in the Turkish Construction Sector [thesis]. İstanbul: İstanbul Technical University; 2006

[2] WEKA. Machine Learning Group at the University of Waikato [Internet]. 1993. Available from: https://www.cs.waikato.ac.nz/ml/index.html [Accessed: 2017-12-05]

[3] KEEL. Knowledge Extraction based on Evolutionary Learning [Internet]. 2005. Available from: http://www.keel.es/ [Accessed: 2017-12-05]

[4] Kaya M, Keleş AE, Laptali Oral E. Construction crew productivity prediction by using data mining methods. Procedia—Social and Behavioral Sciences. 2013;**141**:1249-1253. DOI: 10.1016/j.sbspro.2014.05.215

[5] Andaç MS, Oral EL. Yapım İşlerinde Çalışan Verimliliğinin Yapay Arı Kolonisi Algoritması Kullanılarak Tahmini. In: 3. Proje ve Yapım Yönetimi Kongresi (PYYK 2014); November 06-08, 2014. Antalya. 2014. p. 111

[6] Keleş AE, Kaya M. The Analysis of the Factors Affecting the Productivity in the Wall Construction of the Using Apriori Data Mining Method. In: Proceedings of the XVI. Academic Information Conference (AB2014). Mersin; 05-07 February 2014. p. 831-836

[7] Keleş AE. İnşaat Projelerinde Şantiye Şeflerinin Liderliği ve Çalışan Motivasyonu İlişkisinin Veri Madenciliği ile Belirlenmesi [thesis]. Adana: Çukurova University Natural and Applied Sciences Institute; 2016

[8] Bass BM, Avolio BJ. Mind Garden Tools for Positive Transformation, Multifactor Leadership Questionnaire [Internet]. 1995. Available from: https://www.mindgarden.com/16-multifactor-leadership-questionnaire [Accessed: 2017-12-05]

[9] Dener M, Dörterler M, Orman A. Open Source Data Mining Programs: A Case Study on Weka. In: Proceedings of the XI. Academic Information Conference (AB09). Şanlıurfa; 11-13 February 2009. p. 11-13

[10] Han J, Kamber M. Data Mining: Concept and Techniques. 3rd ed. CA: Academic Press; 2001

[11] Garner SR. Weka: The Waikato environment for knowledge analysis. In: Proceedings of the New Zealand Computer Science Research Students Conference (NZCSRSC). New Zealand; 18-21 April 1995. p. 57-64

[12] Bilgin TT. Data Flow Diagrams Based Data Mining Tools and Software Development Environments. In: Proceedings of the XI. Academic Information Conference (AB09). Şanlıurfa; 11-13 February 2009. p. 807-814

IntechOpen

IntechOpen