Data Mining Approach in Personnel Selection: The Case of the IT Department

Ezgi Demir, Rana Elif Dinçer, Batuhan Atasoy, and Sait Erdal Dinçer

Abstract Data mining studies have been frequently included in the literature recently. Data mining can be applied in every field, especially in banking, marketing, customer relationship management, investment and portfolio management. In the literature, the problem of personnel selection has been examined with the help of multi-criteria decision-making techniques. In this study, it has been aimed to apply data mining techniques in the field of human resources where relatively little has been used. The features of a large-scale construction company have been determined according to the competencies specified in the information technologies department announcement. The candidates were ranked according to these attributes. While ranking, accuracy values have been compared by using basic algorithms of data mining techniques. While applying the process steps, the necessary data pre-processing techniques have been applied to candidates who entered incomplete or incorrect information during the application process. Basically, the decision trees algorithm gave the highest accuracy. Also, random forest, adaboost, gradient boosting, and xgboost algorithms have been tried. In addition, it has found the attributes that should be looked at first in the application features. The high number of data enabled machine learning to learn information more easily and to weigh the existing criteria easily. With this study, it has been aimed to obtain a more objective result by weighting with machine learning algorithms instead of weighting the personnel selection problem with multi-criteria decision-making methodology.

E. Demir (✉)
Department of Management Information Systems, Piri Reis University, Tuzla, Turkey

R. E. Dinçer
Department of Business Informatics, Marmara University, Göztepe, Turkey

B. Atasoy
Department of Mechanical Engineering, Piri Reis University, Tuzla, Turkey
e-mail: batasoy@pirireis.edu.tr

S. E. Dinçer
Department of Econometrics, Marmara University, Göztepe, Turkey
e-mail: edincer@marmara.edu.tr

M. K. Terzioğlu (ed.), *Advances in Econometrics, Operational Research, Data Science and Actuarial Studies*, Contributions to Economics,
https://doi.org/10.1007/978-3-030-85254-2_22

In addition, it is an extremely difficult process to interview candidates for recruitment under the current Covid-19 pandemic conditions that the whole world and our country are struggling with. Online conversations take a lot of time. With this study, it has been aimed to provide optimization by automating the process by weighting the features related to the existing data in the process. The study has been done in the WEKA and Python program.

Keywords Personnel Selection · Data Mining · Information Technologies Department · Artificial Intelligence · Boosting Algorithms

1 Introduction

In today's information society, analyzing data in a meaningful way is a vital issue for businesses. It is an important development to be able to record every transaction made, to create important and meaningful data, and to analyze this data. The large amount of data obtained with the help of information and communication technologies provides convenience to companies, countries, institutions, and organizations in various fields. It has also created various new requirements. In order to adapt to this developing new generation business, understanding data mining methods have been used to discover hidden information in the data. Data mining methods now have been actively used in almost every department of businesses. One of these departments is the human resources, which deals with the manpower and is the main resource of the company. Data mining can be used in the human resources department in areas such as performance management, personnel planning, analysis of various risks, analysis and reporting of important performance criteria. It is also an important tool that can be used to modernize and digitize business processes. In the human resources process, it is extremely important to evaluate the personnel selection according to the qualifications of the personnel and to select the right personnel. A large number of applications have reached the departments. Increasing unemployment processes and decreasing employment rate create increases in applications for positions. Numerous transactions are carried out regarding personnel in human resources departments. Undoubtedly, one of the longest processes is the selection of the personnel to be placed in a position and the negotiation process. In fact, these interviews can continue in 3–4 stages and multiple interviews can be made. All of these processes keep the human resources department busy and prevent them from concentrating on other jobs. This situation reduces the efficiency of the department. At the same time, the process of recruiting personnel requires coordinated work with the managers of the departments where the personnel will be recruited. In this covid-19 pandemic, etc., processes, recruitment and management of new employees in crisis situations have become even more difficult. Staff interviews have been conducted online. Until today, solutions have been found for personnel selection practices in human resources with multi-criteria decision-making methodologies. However, multi-criteria decision-making methodology has judgments that vary from

person to person. With this study, data mining algorithms were used by bringing a new perspective to personnel selection. With machine learning techniques, one of the data mining techniques, the feature values of the applicants were extracted and the personnel selected according to these criteria were made. With this study conducted in the Covid-19 process, a more objective study has been tried to be put forward, away from the interview with candidates. With this study, a model proposal has been presented for the information technology department, which meets the demands of the enterprise in objective conditions in personnel selection. Personnel selection for the information technology department requires a large number of qualities. Multi-criteria decision-making methodology has been used for situations where the number of criteria are high. However, when the number of criteria are high, it becomes difficult to decompose decision criteria that take values from 0–1. In this study, unlike other studies, personnel selection was made with a data mining methodology, which provides more objective results.

This study was conducted within the scope of the application process for the announcement of a holding established in Istanbul in 1987 with code BİM-SRM. In this announcement, it has been planning to hire an information processing officer for the mining department. Seventy-two people applied for the announcement. The criteria required in the application listed have been as follows. Information with 26 criteria was requested. These are application date, announcement code, position, department_name, name, surname, scope of the candidate, candidate's hometown, candidate's age, candidate's gender, candidate's nationality, candidate's military status information, whether the candidate has a travel impediment, the candidate's employment barrier, the candidate's marital status, the name of the university the candidate graduated from, the university department the candidate graduated from, the candidate's starting year, the candidate's license completion year, the number of foreign languages the candidate knows and which ones (English, German, Russian results have been obtained), MsOffice Information on the programs, the candidate's programming language knowledge (results such as Java, C++ were obtained.), the candidate's database information (answers such as Mysql, Oracle, Rdbms), the candidate's knowledge of SAP Modules, and finally, the candidate's work experience on a yearly basis were obtained.

Candidates have been classified according to their criterion degrees according to data mining methodologies and suitable candidates have been determined. Thus, each candidate was not interviewed separately. During the Covid-19 pandemic process, it has been determined that it is the right method to carry out the process with data mining methodologies for recruitment process. Instead of determining the criteria to be considered in the selection of candidates by human resources employees, data mining methodology has determined the weights of the criteria with learning algorithms and classified candidates with high accuracy. As a result of the study, if deemed necessary, the company has the opportunity to choose by making online interviews with the suitable candidates.

2 Modern Approaches in Human Resources Management

Throughout history, the concepts of human and resource management come to the fore wherever there is human and organized labor. From this point of view, it can be understood that this concept is not only an understanding developed as a result of management theories, but a concept that emerged in ancient times and as a result of natural processes (Griffin, 2006). However, as in every field, some updates have been needed in the field of human resources management in order to talk about an effective operation in line with the twenty-first century. There are factors such as the rapid progress in the field of information technologies in the last 20 years, thus the economic globalization has reached extraordinary dimensions and the competitive environment has become global as a result of the limits eliminated by this globalization. Therefore, it has today made it compulsory for both individual employees and organizations more generally to modernize themselves. The concept of owning, holding, developing, and using intellectual capital, which has become a critical factor of superiority in such a global competitive environment, is largely linked to the effectiveness of human resources management. For this reason, modern human resources management approaches have become indispensable for organizations today.

3 Data Mining Application Areas in Organizations

In the growing information society, almost all obstacles to information exchange have been removed. Global knowledge has revealed another dimension of global competition. This developing global environment not only offers a renewed competition environment in terms of workforce quality and quantity. At the same time, data have been gathered from a global environment and evolved into a form that can be compiled. This has led to an increase in the importance given to data and the race to obtain meaningful and useful information from complex data. In such a large and complex information pool, people have come to need forward-looking systems that are difficult to predict as well as internal and external information in order to gain a competitive advantage. Data warehouse, information management, and data mining, which are the three main areas that have developed in response to these needs, are generally aimed at getting more information from data (Silahdaroğlu, 2013). Data mining technology supports companies in finding useful information between large-scale data and mining information. The greatest convenience that data mining technology offers companies is that it helps identify similar trends and patterns of behavior across data sets. This feature is widely used in marketing activities, especially for target markets. Another great help is that it makes relationships that were not visible at first easily visible. For example, a company can analyze the products it sells, design its future campaigns accordingly, discover the links between the products it sells, and develop a marketing tactic based on those links. In today's dynamic business world, there is a high risk that people's decisions will be wrong

or that decision-makers' information will become obsolete. And of course, there is no space for errors in an area where such fragile and momentary changes occur. The only way to reduce such risks is to use decision support solutions that provide knowledge-based solutions. The meaningful data obtained through data mining techniques helps organizations to make strategic decisions correctly, to better manage their risks, and to be innovative. Due to such driving forces, data mining technology has become one of the most important requirements for organizations today.

4 Methodology and Application

In this study, a data mining study was carried out on the CVs of 72 candidates sent to the human resources department for the IT staff to be employed in the mining department. In this context, 5 candidates out of 72 candidates provided incomplete information and were removed from the system during data processing. The remaining 67 candidates were interviewed with human resources personnel, and the candidates' compliance with the starting date requested by the company and their communication skills were evaluated, taking into account their compliance with the company policy and company ownership. Being able to perform data mining applications in human resources does not fall within the scope of an application that can only be done on data. In addition to the qualifications of the individuals, the pre-interview data are an important result in classifying the candidates' data.

In this study, the results of candidates' starting work as a result of the interview with the human resources officer were coded as "suitable" and "not suitable". Candidates' eligibility status has been classified with data mining techniques. Figure 1 shows the data mining application steps. In this context, the needs of the enterprise were analyzed first. As a result of the online interviews, it has been observed that many interviews are held in the human resources process in enterprises, these processes both create a long waiting period for candidates who want to work, depending on the intensity of candidate application, and create an intense loss of work and workforce in the recruitment process for human resources personnel. In the human resources process, the candidates' suitability for the job is evaluated first, and face-to-face interviews are held in the last step, and the suitability of the candidates is evaluated in accordance with the starting date of the company. Thanks to the model, all candidates have been determined as "suitable" and "not suitable", thanks to the determination of their starting and communication skills rather than wasting time while qualitatively separating the candidates, and the candidates and their attributes (attributes) are taught to the machine with the machine learning algorithms used in data mining. "suitable" or "not suitable" will be separated.

Adhering to the data mining application steps, after the needs of the enterprise were understood, the data of the enterprise was taken in the excel file during the applications, and the candidates with the missing data in the data were eliminated in order to prepare for the system. In this context, incompatibilities of the data against the English character have been corrected in order to be loaded into the system. In the

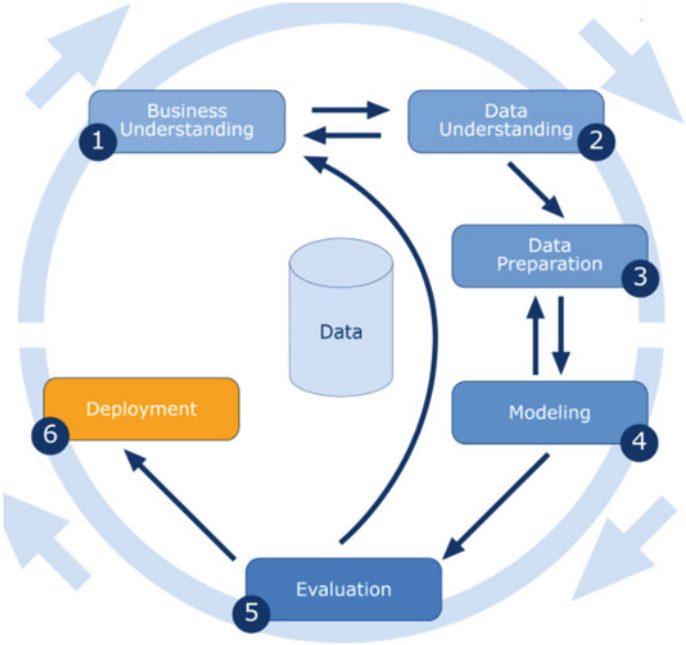


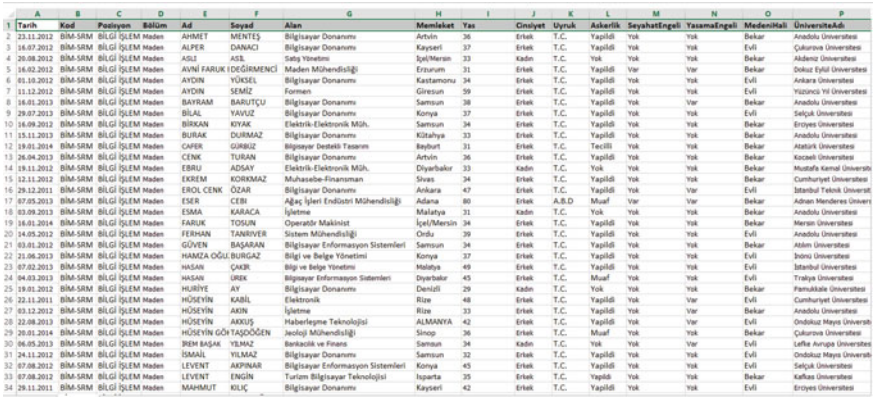
Fig. 1 Data mining implementation steps

data on the military status of the candidates, exemption, postponed, made, and absent (for female candidates) are disaggregated depending on gender. While the number of foreign languages known by the candidates shows a value of 0,1,2, the value 0,1 in the MSOffice programs information is indicated as the candidate's knowledge of the programs: 1, and the status of not knowing: 0. Likewise, the candidate's programming knowledge (Java, C++), database knowledge, and sap modular knowledge are coded as "1", while the candidate's ignorance is coded as "0". Another data editing process is numerical quantities such as 0,3,10,13,14 in the experience period of the candidates. Candidates' work experience increases as the numerical quantity increases in the year column (expressed as the attribute value), their knowledge and experience increase. After the relevant data preparation was defined, when the model was set up during the modeling phase, when the 26 different attribute values of the candidates were examined, it was observed that the application dates, application code, position, department, country, gender information did not affect the results of the candidates', "suitable" or "not suitable". Another factor that makes a difference in the model is that it is seen that the marital status of the candidates does not affect the result, but only the travel disability attribute, so there is no need to use the marital status variable in the model. Because there is a travel disability variable overlapping with the marital status variable. Another factor that makes a difference is that the year of starting the candidates' undergraduate or related departments does not have an effect on the result as an attribute, but instead, the year of graduation of the candidates

should be included in the model because the graduation year affects the result. Before moving on to the modeling results and evaluations, the cross-sections of the data are shown in Fig. 2. Figure 3 shows the cross-sections of the Data Part 2.

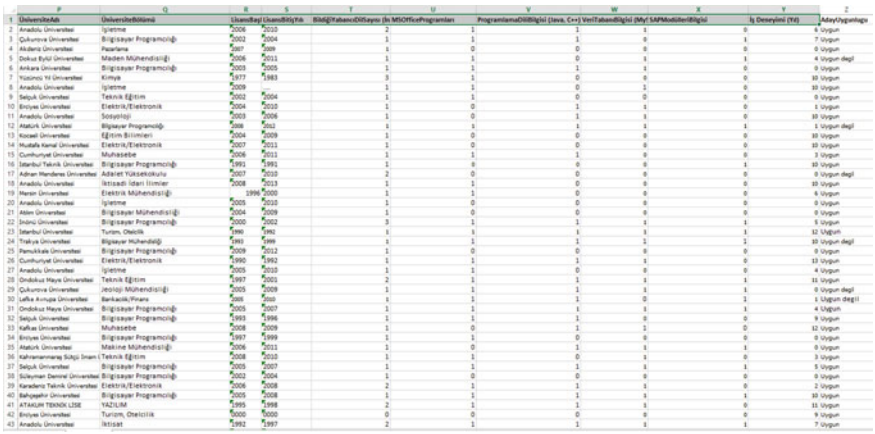
After the data has been modeled, paying attention to the English-Turkish character conformity, it has been converted into csv (comma separated variables) file type and saved. WEKA program is a package program as data mining application techniques. It is extremely easy to use. Figure 4 shows that although the candidates' areas are different, their suitability is homogeneously distributed. Blues, suitable candidates, reds are unsuitable candidates.

In Fig. 4 as can be seen that the blue candidates which come from 23 different areas are appropriate, according to their experiences, the reds are not.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Tarih	Kod	Pozisyon	Bölüm	Ad	Soyad	Alan	Memleket	Yas	Cinsiyet	Uyruk	Askerlik	SeyahatEngeli	YasamaEngeli	MedeniHali	ÜniversiteAdı
2	23.11.2012	BİM-SRM	BİLGİ İŞLEM	Maden	AHMET	MENTEŞ	Bilgisayar Donanımı	Artvin	36	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Anadolu Universitesi
3	16.07.2012	BİM-SRM	BİLGİ İŞLEM	Maden	ALPER	DANACI	Bilgisayar Donanımı	Kayseri	37	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Çukurova Universitesi
4	20.08.2012	BİM-SRM	BİLGİ İŞLEM	Maden	ASLI	AEŞ	Satış Yönetimi	İçel/Mersin	33	Kadın	T.C.	Yok	Yok	Yok	Bekar	Akdeniz Universitesi
5	16.02.2012	BİM-SRM	BİLGİ İŞLEM	Maden	AVNİ FARUK	İDEĞİRMENCİ	Maden Mühendisliği	Erzurum	31	Erkek	T.C.	Yapildi	Var	Var	Bekar	Dokuz Eylül Universitesi
6	01.10.2012	BİM-SRM	BİLGİ İŞLEM	Maden	AYDIN	YÜKSEL	Bilgisayar Donanımı	Kastamonu	34	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Ankara Universitesi
7	11.12.2012	BİM-SRM	BİLGİ İŞLEM	Maden	AYDIN	SEMİZ	Formen	Giresun	59	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Yüzüncü Yıl Universitesi
8	16.01.2013	BİM-SRM	BİLGİ İŞLEM	Maden	BAYRAM	BARUTÇU	Bilgisayar Donanımı	Samsun	38	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Anadolu Universitesi
9	29.07.2013	BİM-SRM	BİLGİ İŞLEM	Maden	BİLAL	YAVUZ	Bilgisayar Donanımı	Konya	37	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Selçuk Universitesi
10	16.09.2012	BİM-SRM	BİLGİ İŞLEM	Maden	BİRKAN	KIYAK	Elektrik-Elektronik Müh.	Samsun	34	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Eroyes Universitesi
11	15.11.2013	BİM-SRM	BİLGİ İŞLEM	Maden	BURAK	DURMAZ	Bilgisayar Donanımı	Kütahya	33	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Anadolu Universitesi
12	19.01.2014	BİM-SRM	BİLGİ İŞLEM	Maden	CAFER	GÜRBÜZ	Bilgisayar Destekli Tasarım	Bayburt	31	Erkek	T.C.	Teciili	Yok	Yok	Bekar	Atatürk Universitesi
13	26.04.2013	BİM-SRM	BİLGİ İŞLEM	Maden	CENK	TURAN	Bilgisayar Donanımı	Artvin	36	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Kocaeli Universitesi
14	19.11.2012	BİM-SRM	BİLGİ İŞLEM	Maden	EBRU	ADSAY	Elektrik-Elektronik Müh.	Diyarbakır	33	Kadın	T.C.	Yok	Yok	Yok	Bekar	Mustafa Kemal Universit
15	12.11.2012	BİM-SRM	BİLGİ İŞLEM	Maden	EKREM	KORKMAZ	Muhasebe-Finansman	Sivas	34	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Cumhuriyet Universitesi
16	29.12.2011	BİM-SRM	BİLGİ İŞLEM	Maden	EROL CENK	ÖZAR	Bilgisayar Donanımı	Ankara	47	Erkek	T.C.	Yapildi	Yok	Var	Evli	İstanbul Teknik Universit
17	07.05.2013	BİM-SRM	BİLGİ İŞLEM	Maden	ESER	CEBİ	Ağaç İşleri Endüstri Mühendisliği	Adana	80	Erkek	A,B,D	Muaf	Var	Var	Bekar	Adnan Menderes Univers
18	03.09.2013	BİM-SRM	BİLGİ İŞLEM	Maden	ESMA	KARACA	İşletme	Malatya	31	Kadın	T.C.	Yok	Yok	Yok	Bekar	Anadolu Universitesi
19	16.01.2014	BİM-SRM	BİLGİ İŞLEM	Maden	FARUK	TOSUN	Operatör Makinist	İçel/Mersin	34	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Mersin Universitesi
20	14.05.2012	BİM-SRM	BİLGİ İŞLEM	Maden	FERHAN	TANRIVER	Sistem Mühendisliği	Ordu	39	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Anadolu Universitesi
21	03.01.2012	BİM-SRM	BİLGİ İŞLEM	Maden	GÜVEN	BAŞARAN	Bilgisayar Enformasyon Sistemleri	Samsun	34	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Atılım Universitesi
22	21.06.2013	BİM-SRM	BİLGİ İŞLEM	Maden	HAMZA OĞLU	BURGAZ	Bilgi ve Belge Yönetimi	Konya	37	Erkek	T.C.	Yapildi	Yok	Yok	Evli	İnönü Universitesi
23	07.02.2013	BİM-SRM	BİLGİ İŞLEM	Maden	HASAN	ÇAKIR	Bilgi ve Belge Yönetimi	Malatya	49	Erkek	T.C.	Yapildi	Yok	Yok	Evli	İstanbul Universitesi
24	04.03.2013	BİM-SRM	BİLGİ İŞLEM	Maden	HASAN	ÜREK	Bilgisayar Enformasyon Sistemleri	Diyarbakır	45	Erkek	T.C.	Muaf	Yok	Yok	Evli	Trakya Universitesi
25	19.01.2012	BİM-SRM	BİLGİ İŞLEM	Maden	HURİYE	AY	Bilgisayar Donanımı	Denizli	29	Kadın	T.C.	Yok	Yok	Yok	Bekar	Pamukkale Universitesi
26	22.11.2011	BİM-SRM	BİLGİ İŞLEM	Maden	HÜSEYİN	KABİL	Elektronik	Rize	48	Erkek	T.C.	Yapildi	Yok	Var	Evli	Cumhuriyet Universitesi
27	03.12.2012	BİM-SRM	BİLGİ İŞLEM	Maden	HÜSEYİN	AKIN	İşletme	Rize	33	Erkek	T.C.	Yapildi	Yok	Var	Bekar	Anadolu Universitesi
28	22.08.2013	BİM-SRM	BİLGİ İŞLEM	Maden	HÜSEYİN	AKKUŞ	Haberleşme Teknolojisi	ALMANYA	42	Erkek	T.C.	Yapildi	Yok	Var	Evli	Ondokuz Mayıs Universit
29	20.01.2014	BİM-SRM	BİLGİ İŞLEM	Maden	HÜSEYİN GÖK	TAŞDÖĞEN	Jeoloji Mühendisliği	Sinop	36	Erkek	T.C.	Muaf	Yok	Yok	Bekar	Çukurova Universitesi
30	06.05.2013	BİM-SRM	BİLGİ İŞLEM	Maden	İREM BAŞAK	YILMAZ	Bankacılık ve Finans	Samsun	34	Kadın	T.C.	Yok	Yok	Var	Evli	Lefke Avrupa Universites
31	24.11.2012	BİM-SRM	BİLGİ İŞLEM	Maden	İSMAİL	YILMAZ	Bilgisayar Donanımı	Samsun	32	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Ondokuz Mayıs Universit
32	07.08.2012	BİM-SRM	BİLGİ İŞLEM	Maden	LEVENT	AKPINAR	Bilgisayar Enformasyon Sistemleri	Konya	45	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Selçuk Universitesi
33	07.08.2012	BİM-SRM	BİLGİ İŞLEM	Maden	LEVENT	ENGİN	Turizm Bilgisayar Teknolojisi	Isparta	35	Erkek	T.C.	Yapildi	Yok	Yok	Bekar	Kafkas Universitesi
34	29.11.2011	BİM-SRM	BİLGİ İŞLEM	Maden	MAHMUT	KILIÇ	Bilgisayar Donanımı	Kayseri	42	Erkek	T.C.	Yapildi	Yok	Yok	Evli	Eroyes Universitesi

Fig. 2 Cross-sections of the data



	P	Q	R	S	T	U	V	W	X	Y	Z
1	ÜniversiteAdı	ÜniversiteBölümü	LisansBaşl	LisansBitiş	BildiğiYabancıDilSayısı	İlk MSOfficeProgramları	ProgramlamaDilBilgisi (Java, C++)	VeriTabanıBilgisi	(My' SAPModülleriBilgisi	İş Deneyimi (Yıl)	AdayUygunlugu
2	Anadolu Universitesi	İşletme	2006	2010	2	1	1	1	0	4	Uygun
3	Çukurova Universitesi	Bilgisayar Programcılığı	2002	2004	1	1	1	0	1	7	Uygun
4	Akdeniz Universitesi	Pazarlama	2007	2009	1	0	0	0	0	0	Uygun
5	Dokuz Eylül Universitesi	Maden Mühendisliği	2006	2011	1	1	0	1	1	4	Uygun degil
6	Ankara Universitesi	Bilgisayar Programcılığı	2003	2005	1	1	1	1	0	0	Uygun
7	Yüzüncü Yıl Universitesi	Kimya	1977	1983	3	1	0	0	0	10	Uygun
8	Anadolu Universitesi	İşletme	2009		1	1	0	1	0	10	Uygun
9	Selçuk Universitesi	Teknik Eğitim	2002	2004	1	1	0	0	0	0	Uygun
10	Eroyes Universitesi	Elektrik/Elektronik	2004	2010	1	0	1	1	0	1	Uygun
11	Anadolu Universitesi	Sosyoloji	2003	2006	1	0	1	1	0	10	Uygun
12	Atatürk Universitesi	Bilgisayar Programcılığı	2008	2012	1	1	1	1	1	1	Uygun degil
13	Kocaeli Universitesi	Eğitim Bilimleri	2004	2009	1	0	0	0	0	10	Uygun
14	Mustafa Kemal Universitesi	Elektrik/Elektronik	2007	2011	1	0	0	0	0	10	Uygun
15	Cumhuriyet Universitesi	Muhasebe	2006	2011	1	1	1	1	0	3	Uygun
16	İstanbul Teknik Universitesi	Bilgisayar Programcılığı	1991	1991	1	1	0	1	0	10	Uygun
17	Adnan Menderes Universitesi	Adalet Yüksekokulu	2007	2010	2	0	0	0	0	0	Uygun degil
18	Anadolu Universitesi	İktisadi İdari İlimler	2008	2013	1	1	0	0	0	10	Uygun
19	Mersin Universitesi	Elektrik Mühendisliği	1996	2000	1	1	1	0	0	6	Uygun
20	Anadolu Universitesi	İşletme	2005	2010	1	0	0	0	0	0	Uygun
21	Atılım Universitesi	Bilgisayar Mühendisliği	2004	2009	1	0	0	0	0	0	Uygun
22	İnönü Universitesi	Bilgisayar Programcılığı	2000	2002	3	1	1	1	1	5	Uygun
23	İstanbul Universitesi	Turizm, Otelcilik	1990	1992	1	1	1	1	1	12	Uygun
24	Trakya Universitesi	Bilgisayar Mühendisliği	1993	1999	1	1	1	1	1	10	Uygun degil
25	Pamukkale Universitesi	Bilgisayar Programcılığı	2009	2012	1	0	0	0	0	0	Uygun
26	Cumhuriyet Universitesi	Elektrik/Elektronik	1990	1992	1	1	1	1	0	13	Uygun
27	Anadolu Universitesi	İşletme	2005	2010	1	1	0	0	0	4	Uygun
28	Ondokuz Mayıs Universitesi	Teknik Eğitim	1997	2001	2	1	1	1	1	11	Uygun
29	Çukurova Universitesi	Jeoloji Mühendisliği	2003	2009	1	1	1	1	0	0	Uygun degil
30	Lefke Avrupa Universitesi	Bankacılık/Finans	2005	2010	1	1	1	0	1	1	Uygun degil
31	Ondokuz Mayıs Universitesi	Bilgisayar Programcılığı	2005	2007	1	1	1	1	1	4	Uygun
32	Selçuk Universitesi	Bilgisayar Programcılığı	1993	1996	1	1	1	1	0	9	Uygun
33	Kafkas Universitesi	Muhasebe	2008	2009	1	0	1	1	0	12	Uygun
34	Eroyes Universitesi	Bilgisayar Programcılığı	1997	1999	1	1	0	0	0	0	Uygun
35	Atatürk Universitesi	Makine Mühendisliği	2006	2011	1	0	0	1	0	0	Uygun
36	Kahramanmaraş Sütçü İmam Ü	Teknik Eğitim	2008	2010	1	1	0	1	0	3	Uygun
37	Selçuk Universitesi	Bilgisayar Programcılığı	2005	2007	1	1	1	1	1	5	Uygun
38	Süleyman Demirel Universitesi	Bilgisayar Programcılığı	2002	2004	1	0	0	0	0	0	Uygun
39	Karadeniz Teknik Universitesi	Elektrik/Elektronik	2006	2008	2	1	1	1	0	2	Uygun
40	Bahçeşehir Universitesi	Bilgisayar Programcılığı	2005	2008	1	1	1	1	1	10	Uygun
41	ATAKUM TEKNİK LİSE	YAZILIM	1995	1998	2	1	1	1	0	11	Uygun
42	Eroyes Universitesi	Turizm, Otelcilik	2000	2000	0	0	0	0	0	9	Uygun
43	Anadolu Universitesi	İktisat	1992	1997	2	1	1	1	1	7	Uygun

Fig. 3 Cross-sections of the data part 2

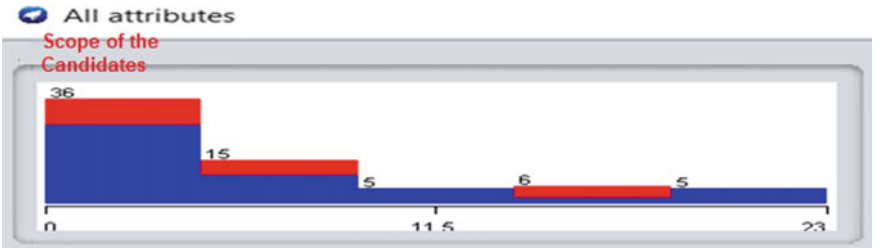


Fig. 4 Distribution of the candidates' fields

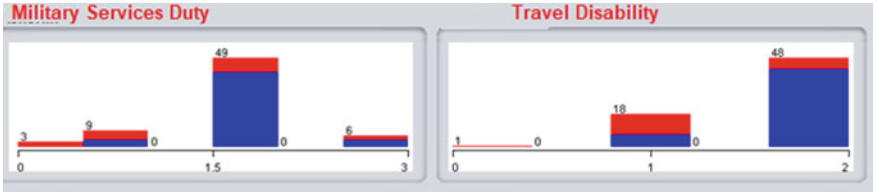


Fig. 5 Distribution of "military service" and "travel disability" status of the candidates

In Fig. 5, the suitability of the candidates according to their military service status and travel disability has been examined. In the case of military service, it is seen that the most suitable ones from four categories (done, postponed, exempted-not, and not) have done their military service, then exempted-no, postponed, and in the last case those who did not.

Figure 5 also looks at the travel disability status of the candidates, and it is seen that the blue density of the candidates who do not have travel disability is more appropriate in the current two situations (there is a disability, there is no disability). However, it cannot be said that every candidate without travel disability is eligible, as there are red "unsuitable" candidates in the histogram on the far right without travel barriers. This shows that the data is distributed homogeneously.

In Fig. 6, the candidates' suitability according to the name-departments of the university has been examined.

As can be seen in Fig. 6, the eligibility of the candidates according to the university names and departments is homogeneously distributed. Blue indicates conformity, red indicates unfit. Due to the high number of applications from computer, hardware,

Fig. 6 University names and departments of candidates

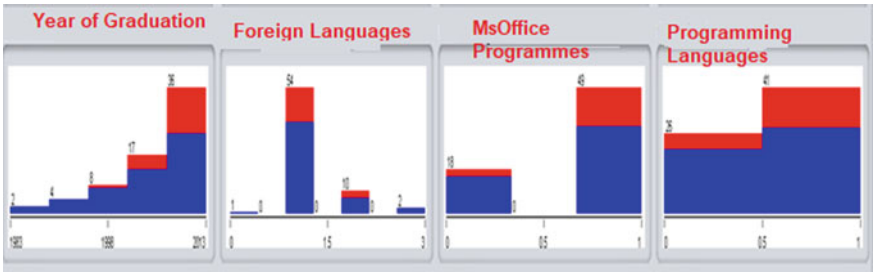


Fig. 7 Licensing expiry year, number of foreign languages, office program, and programming information of candidates

and information departments, the number of eligible candidates on the far right side of the University Department chart is as high, while there are many non-eligible candidates from the same department due to the high competition for this application announcement.

In Fig. 7, license expiry year, number of foreign languages, office program, and programming information of the candidates are taken into consideration. As the graduation years of the candidates approached today, their eligibility increased, and the unsuitability status due to their recent graduates and their inexperience increased. There are four categories in the number of foreign languages which are known. There is a suitable candidate who does not speak (0: no spoken, 1: speak one language, 2: speak 2 languages, 3: speak 3 languages). While the majority of suitable candidates who know one language are in the majority, most of those who are not suitable are those who speak one language.

In Fig. 7, there are two categories as those who know office programs and those who do not. The most suitable ones are seen in the category of those who know. Likewise, those who know java and c++ have also made a significant difference in compliance. In Fig. 8, the candidates' database information, sap modular knowledge, work experience, and candidate eligibility in the last case have been examined.

According to what is given in Fig. 8, the suitability of those who know the database has increased by far, the suitability of those with sap modular knowledge increases by far. The eligibility of those with more experience in work experience has increased.

Fig. 8 Candidates' database information, sap modular knowledge, work experience, and candidate eligibility in the last case

The eligibility of newly graduated candidates is as high. Data are homogeneously distributed. In the last case, 49 of the candidates are eligible, while 18 of them are not.

Prediction of candidates was made by applying data mining. It was modeled in the python program.

The decision tree algorithm has been selected and 80% of the data have been parsed as learning data, while 20% of the data have been separated as test data.

The results of running the data have been shown in Fig. 9. Relation part shows the value where the file name is registered. Instances shows the number of rows (number of people) to be analyzed. The attribute values show the feature values of the lines. Here, candidates have 20 properties. In machine learning, learning algorithms have been divided into either 67% training, 33% test data or 80% training, and 20% test data. A higher learning rate always gives better results. Therefore, 80% learning is reserved as 20% test data. Train 80% expression is the learning rate. Remain (the rest) is test data. And also the model has been tried for random forest algorithm. The model results have been shown in Fig. 10. Another model has been tried for Adaboost algorithm in Fig. 11. Gradient boosting and Xgboost algorithms have been tried in Figs. 12 and 13, respectively.

The accuracy of all the models has been shown in Fig. 14. Accuracy percentages have been given according to the cross validation evaluations of the models. Best results have been seen when cross validation is 4 and 8 in Fig. 14.

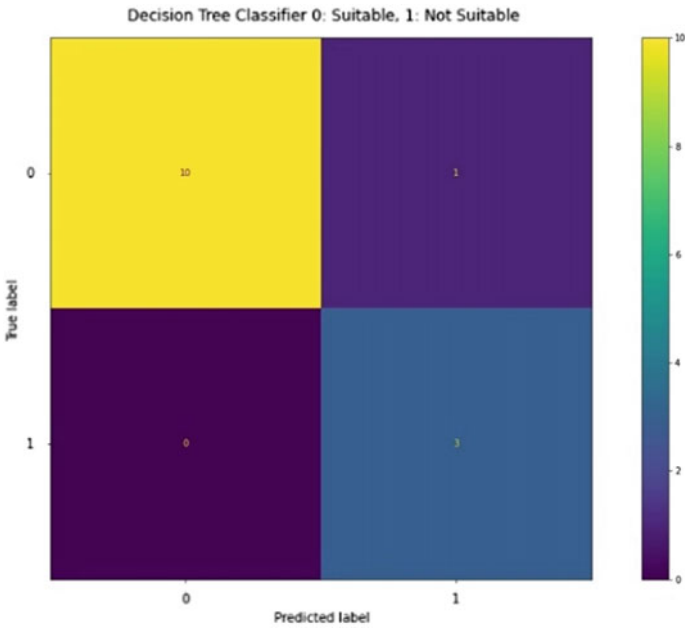


Fig. 9 Results of the data for decision tree algorithm

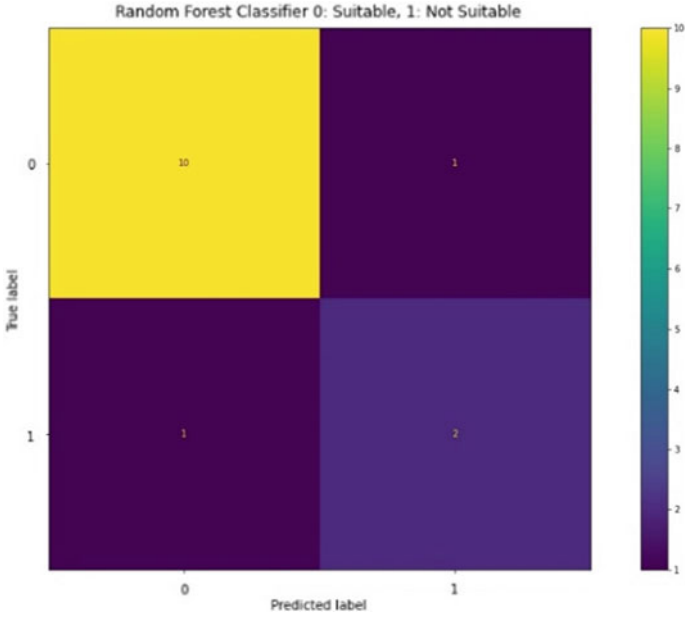


Fig. 10 Results of the model for random forest algorithm

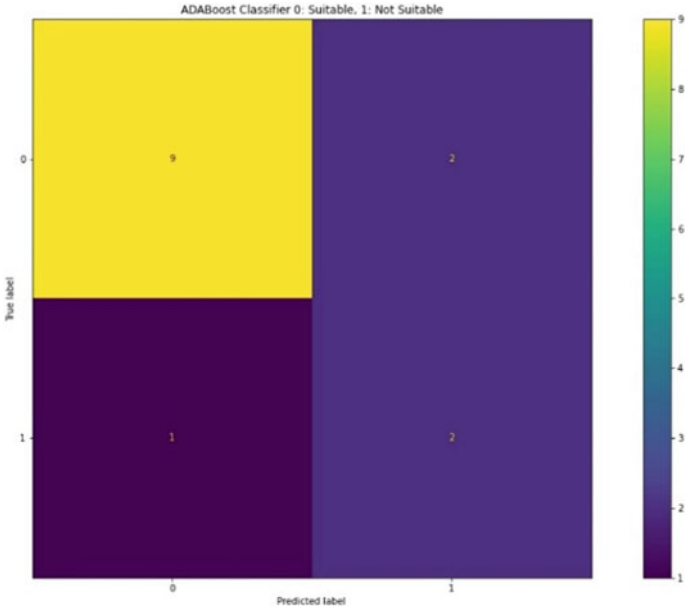


Fig. 11 Results of the model for ADABoost algorithm

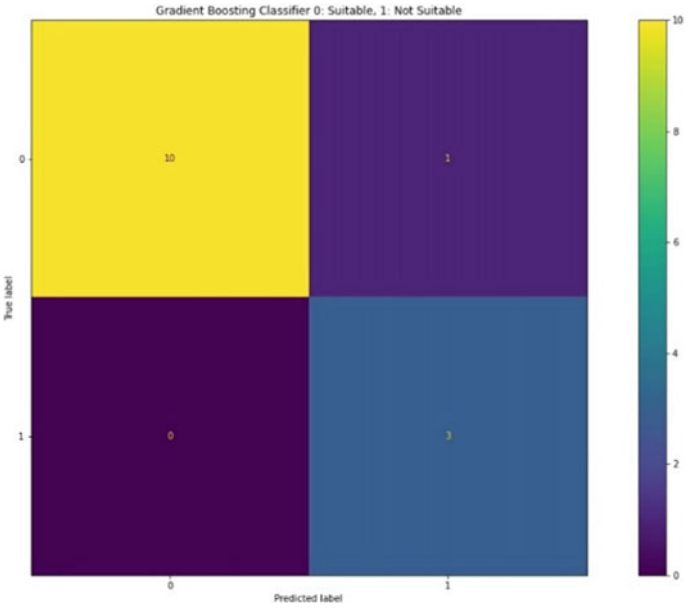


Fig. 12 Results of the model for Gradient Boosting algorithm

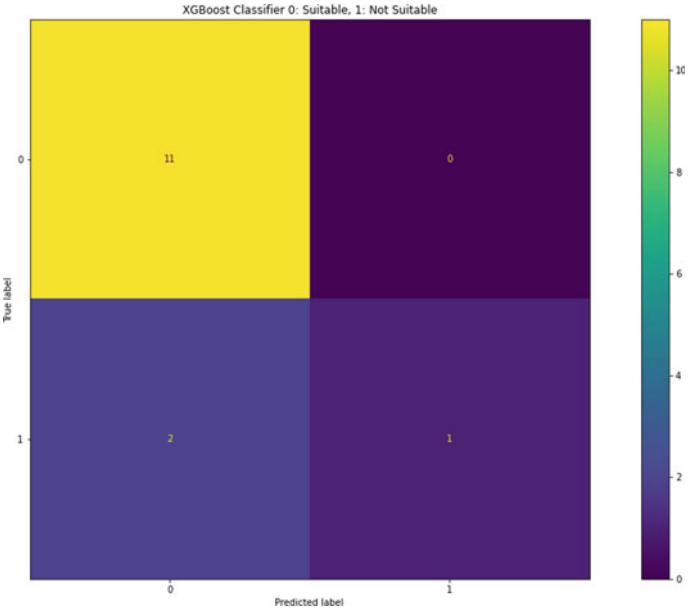


Fig. 13 Results of the model for XGBoost algorithm

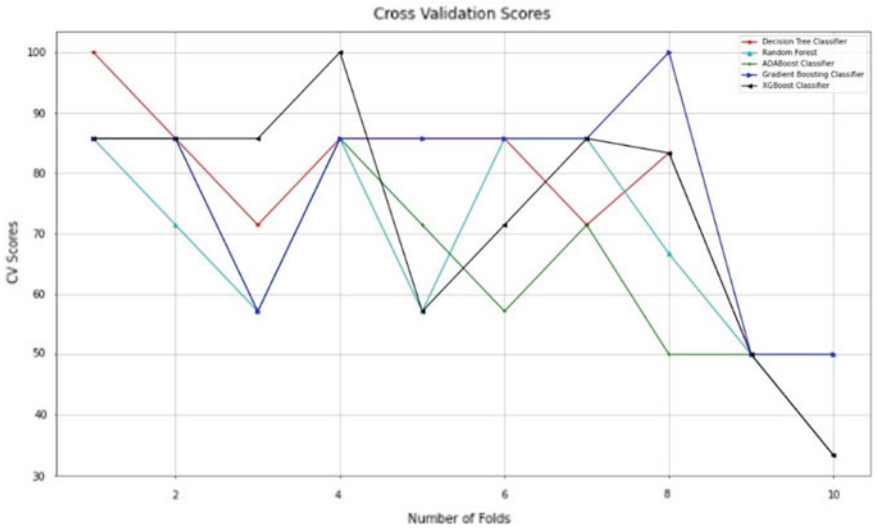


Fig. 14 Cross validation scores

5 Conclusion

Data mining is an important step in the process of finding information, which is known as Information and Data Discovery in the literature, which reveals data with various methods (Şeker, 2016). Data Mining is a research and iterative process that separates and filters the patterns in the knowledge discovery process and prepares them for the next step (Coşlu, 2013).

When examined in detail, it can be seen that the data mining process includes the following steps (Akpınar, 2000):

1. Creating a data set or obtaining it from a ready source,
2. Organizing the data (minimizing repetition, filling missing parts in the data, etc.),
3. Selecting the appropriate data mining technique and algorithm for the project,
4. Application and interpretation of the model,
5. Reporting and presenting the obtained information.

The first thing that matters in a data mining application is to set a roadmap based on the goals and framework of the project. In this study, the recruitment processes of a construction company during the pandemic process were examined. During the pandemic process, it is extremely difficult to have long and multiple interviews with candidates. For this reason, employee selection was made with data mining techniques. First, the characteristics of the candidates for recruitment were determined. Later, the candidates were examined individually according to their eligibility. Finally, by applying the machine learning methodology, a hierarchical structure was

created according to the criteria that should be considered in the recruitment process of the candidates. Candidates who were eligible for recruitment were classified.

With this study, it is aimed to gain a different methodology to the subject of personnel selection than the multi-criteria decision-making methodology, which is frequently used in the literature. Expert opinions and quantitative value judgments play an important role in multi-criteria decision-making methodology. In data mining and machine learning techniques, more objective results can be obtained as the criteria are determined by the machine. In this respect, the study is aimed to shed light on multi-criteria and data science studies in the literature.

References

Akpınar H (2000) Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. İşletme Fakültesi Dergisi, 1–22

Coşlu E (2013) Veri Madenciliği. Akdeniz Üniversitesi, Antalya: XV. Akademik Bilişim Konferansı Bildiriler

Griffin R (2006) Principles of management. Houghton Mifflin Company

Silahdaroğlu G (2013) Veri Madenciliği . İstanbul: Papatya Yayınları

Şeker E (2016) Veri Madenciliği Araçları YBS Ansiklopedi Cilt 3, Sayı 4. YBS Ansiklopedi Cilt 3, Sayı 4