SDÜ İKY Dergisi
2023, 2(2), ss. 44-53
Araştırma Makalesi

e-ISSN: 2979-997X
https://dergipark.org.tr/tr/pub/sduikyd

SDU Journal of HRM
2023, 2(2), pp. 44-53
Research Article

Başvuru: 06.11.2023 **Kabul:** 26.12.2023

A Data Mining Application for Classification of Performance Values Under the Conditions of Digitalization of HR Processes*

İK Süreçlerinin Dijitalleşmesi Koşullarında Performans Değerlerinin Sınıflandırılmasına Yönelik Veri Madenciliği Uygulaması

Gözde KATIRCIOĞLU[1]
Emel KIZILKAYA AYDOĞAN[2]
Yılmaz DELİCE[3]

Abstract

Human resources management is the key to maintaining and developing quality personnel in modern companies. Analyzing the human resources data of institutions, identifying problems, and determining the strategy provide a significant competitive advantage. The topic of digitalization of human resources management in businesses is increasing day by day. With the fourth industrial revolution, also known as Industry 4.0, digitalization also affects human resources management (HRM), which is seen as the basis of corporate business and transactions. In this field, more and more businesses are investing in human resources management, which increases efficiency and savings with artificial intelligence and data mining methods. Data mining applications are used to make sense of data and transform it into information and to assist organizations in decision-making processes. The aim of the present study is to examine the impact of digitalization on human resources management practices and to identify trends in this regard. In this context, decision tree and rule extraction algorithms, which are data mining methods, were used. Experiments and comparison studies conducted on real data have revealed that it is quite good at categorizing the available data. The rules obtained enabled the company to have insight into the behavior of potential employee candidates. The importance of this study is to reveal the positive effects of digitalization and artificial intelligence in the business world on human resources management with the developed technique.

Keywords: *Human Resources Management (HRM), Data Mining, Digitalization, Decision Tree*
Jel Codes: *M12, O15, O33*

Öz

İnsan kaynakları yönetimi, modern şirketlerde kaliteli personeli korumanın ve geliştirmenin anahtarıdır. Kurumların insan kaynakları verilerinin analiz edilmesi, sorunların tespit edilmesi ve stratejinin belirlenmesi avantaj sağlamaktadır. İşletmelerde insan kaynakları yönetiminin (İKY) dijitalleşmesi konusu her geçen gün artmaktadır. Endüstri 4.0 olarak da bilinen dördüncü sanayi devrimiyle birlikte dijitalleşme, kurumsal iş ve işlemlerin temeli olarak görülen insan kaynakları yönetimini de etkilemektedir. Bu alanda giderek daha fazla işletme, yapay zeka ve veri madenciliği yöntemleriyle verimliliği ve tasarrufu artıran insan kaynakları yönetimine yatırım yapmaktadır. Veri madenciliği uygulamaları ise, verileri anlamlandırıp bilgiye dönüştürmek ve kuruluşlara karar verme süreçlerinde yardımcı olmak amacıyla kullanılmaktadır. Bu çalışmanın amacı dijitalleşmenin insan kaynakları yönetimi uygulamaları üzerindeki etkisini incelemek ve bu konudaki eğilimleri tespit etmektir. Bu kapsamda veri madenciliği yöntemlerinden karar ağacı ve kural çıkarma algoritmaları kullanılmıştır. Gerçek veriler üzerinde yapılan deneyler ve karşılaştırma çalışmaları, mevcut verileri kategorize etme konusunda oldukça iyi olduğunu ortaya koymuştur. Elde edilen kurallar, şirketin potansiyel çalışan adaylarının davranışları hakkında fikir sahibi olmasını sağlamıştır. Bu çalışmanın önemi, iş dünyasında dijitalleşmenin ve yapay zekanın insan kaynakları yönetimine olumlu etkilerini geliştirilen teknikle ortaya koymaktır.

Anahtar Kelimeler: *İnsan Kaynakları Yönetimi (İKY), Veri Madenciliği, Dijitalleşme, Karar Ağacı*
Jel Kodları: *M12, O15, O33*

Önerilen Atıf / Suggested Citation: Katırcıoğlu, G., Kızılkaya Aydoğan, E., & Delice, Y., (2023). A Data Mining Application for Classification of Performance Values Under the Conditions of Digitalization of HR Processes. *Süleyman Demirel Üniversitesi İnsan Kaynakları Yönetimi Dergisi*, *2*(2), 44-53.

* The study is prepared from the paper titled "A Data Mining Application for Classification of Performance Values Under the Conditions of Digitalization of HR Processes" that was presented on 14 October 2023 in the 1st International Human Resources Management Congress and was not published as full-paper in the Proceedings Book.

[1] Dr., Erciyes University, Faculty of Engineering, Department of Industrial Engineering, gozdekatircioğlu@gmail.com, https://orcid.org/0000-0003-0748-7301
[2] Prof. Dr., Erciyes University, Faculty of Engineering, Department of Industrial Engineering, ekaydogan@erciyes.edu.tr, https://orcid.org/0000-0003-0927-6698
[3] Doç. Dr., Kayseri University, Faculty of Applied Sciences, Department of International Trade and Logistics, ydelice@kayseri.edu.tr, https://orcid.org/0000-0002-4654-0526

Introduction

Nowadays, with the development of technology, Concepts such as information technologies, internet, digitalization, machine learning and artificial intelligence have come to the fore. The process of change in the technological field has affected social life and businesses. Businesses have also had to adapt to this change. Technological transformation and digitalization are the leading developments that shape the business world. Human resources management (HRM) is one of the basic management functions that adapt to this change.

Digital technologies have transformed traditional human resource management processes, the structure and functions of HR departments, the activities of relevant personnel and the entire human capital-based value chain. Digital transformation has created many opportunities for companies to adapt to technological transformation (Ulrich and Dulebohn, 2015). With advanced HR management, businesses have managed their HR processes more effectively, made better recruitment decisions, and made workforce optimization and future planning more effective by predicting employees who are considering leaving their jobs. The coronavirus outbreak and the resulting economic crisis have increased the importance of HRM digitalization. As a result, businesses began to invest in human resources management that increases efficiency and savings with artificial intelligence and data mining methods.

Digitalization, an important element of technological transformation, basically refers to the application of digital technologies to significantly improve business performance indicators such as workforce efficiency and customer service, optimize operations or develop new business models (Vial, 2019). Digitalization makes business life easier by providing easier and faster access to more data and managing the obtained data. Digital HRM is the application of web-based technological solutions to the human resources management activities of organizations (Ruël et al., 2007). According to Bondarouk and Ruël (2009), digital HRM is an inclusive term that aims to create value for employees and management both within the organization and in general and includes all possible integration mechanisms between HRM and information technologies.

The digitalization process is closely related to business intelligence, digital transformation, data mining (De Haes et al., 2020; Bongiorno et al., 2018) and information management (Mulyana et al., 2021). During the digital transformation process, data has become the most important digital asset of companies. However, the data produced by computer systems become meaningful when processed for a specific purpose. For this reason, techniques capable of processing significant amounts of data gain great importance (Kalikov, 2006). Converting and analyzing the large-sized, unprocessed raw data obtained into meaningful form is possible with data mining techniques (Erdem and Özdağoğlu, 2008). Data mining is a process that uses many data analysis tools to reveal hidden information and structures in large databases (Zhou, 2002).

When the studies on this subject are examined in general; In particular, it has been observed that human resources management and digitalization are less discussed. For example, Uğurlu and Doğan (2023) studied the impact of digital HRM transformation and digital developments on the recruitment function. Yılmaz et al (2023) studied how Industry 4.0 can affect HRM processes. Karaboğa et al. (2022) examined and classified the use of digitalization and digital technologies in HRM with bibliometric analysis with the data they obtained from the Web of Science Core Collection database. Akduman (2019) explained the impact of the digital world on the human resources recruitment function and evaluated it with sample applications. In their studies, Filizöz and Orhan (2018) discussed the subject of Industry 4.0, human and service concepts with recent studies. In this present study aims to present a strategy for integrating data mining techniques into the human resources management system. The innovative aspect of the study is to analyze the reasons for leaving the job with data mining methods and present a rule-based decision mechanism to human resources departments. According to personnel data obtained from a production factory, the reasons for dismissal were examined using performance criteria and rule extraction and classification analysis from data mining methods, and a rule-based system was chosen by classifying these reasons. The rules obtained from classification algorithms provide an effective reference and insight for the company's human resources manager to make decisions.

The rest of the paper is organized as follows: Section 2 presents the materials used, data mining, decision tree and decision tree algorithms. Section 3 presents the performance measures used and the findings obtained. Finally, Chapter 4 presents the conclusions of the study.

1. Materials and Methods

1.1. Materials

The empirical basis of the study was provided based on data received from a company producing in the industry sector. The data was conducted in January-December 2022. Personnel data includes personnel collar type, age, educational status, and operation time. A total of 2004 pieces of data were obtained. Missing data was detected at a rate of 2% in the data set, and records containing missing data were cleared from the data set. Personnel collar types are grouped as blue collar and white collar. The age group was chosen 18-60. Educational status was grouped as primary school, high school, junior college and bachelor/master. Operation time of the personnel are calculated in months. There is information about two thousand personnel.

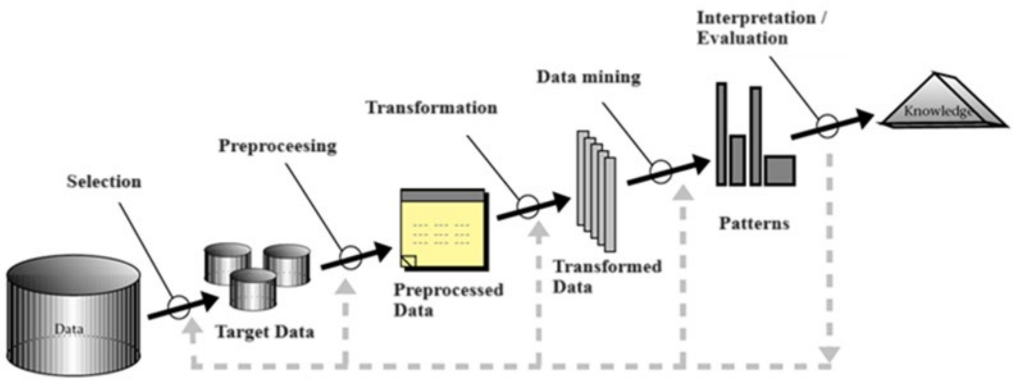
Reasons for leaving the job determined as output parameters are divided into 6 classes: absenteeism, abuse of trust (such as coming to work drunk, endangering the safety of the job, damaging workplace property, fighting), job termination, resignation, military/marriage and retirement.

1.2. Data Mining

During the digital transformation process, data has become the most important digital asset of companies. At the same time, techniques that could process significant amounts of data are also gaining great importance (Kalikov, 2006). Data mining; it is one of the techniques used to reveal previously undiscovered information based on a wide variety and large amount of data held in data warehouses and to use it to make decisions and realize action plans. In addition, data mining is the process of transforming previously unknown, interesting, unusual and potentially useful data in the system into meaningful information (Shen, 2007).

Figure 1

Information Discovery Process in Database



Source: Maimon and Rokach, 2005.

The data mining stages are shown in Figure 1. The first stage of data mining is selecting the data set to be analysed. Selecting the right data is considered the first step in reaching the right conclusion. After the data is selected, the data is cleaned and combined with data sets from different sources, if any. Finally, it is aimed to obtain useful information by categorizing the relationships between the data (Yıldırım et al., 2017).

Data mining models are basically grouped into 3 groups as Classification and Regression, Clustering and Association Rules. Classification is one of the widely used data mining techniques that separates new samples into predefined classes. Regression is used to estimate continuous values while classification is used to estimate categorical values. Clustering is a mining technique in which interesting patterns are discovered from the database (Amershi & Conati, 2006). Association Rules are techniques used to find common patterns, associations, correlations and coincidental structures between sets of objects in moving databases (Chen et al., 2005).

1.2.1. Decision Trees

Decision trees are one of the most widely used classification techniques because they are easy to implement, interpret and integrate (Argüden & Erşahin, 2008). Decision trees based on information theory create models that enable easy interpretation of complex and unknown data. The basic structure of a decision tree consists of three basic parts called node, branch and leaf. In this tree structure, nodes represent each feature, while branches and trees are other elements of the tree structure. The top part is called the root, the parts between the roots and leaves are referred to as branches and the last part of the tree is called the leaf (Quinlan, 1993). Stems show the conditions of the features, while leaves reveal the classification results. The final form of the tree could be translated into a series of If-Then-Else rules.

The most important step in creating decision trees is to create the tree structure according to the attribute values of the branching in the tree. There are various approaches developed in the literature to solve this problem. The most important of these are information gain and information gain ratio (Quinlan, 1993), Gini index (Breiman et al., 1984), Towing rule (Breiman et al., 1984) and Chi-Square probability table statistic (Mingers, 1989) approaches. In the information gain and information gain ratio approach, during the creation of the tree, segmentation is made based on the entropy values of each class label in the training data set. This process is repeated recursively until the repetition becomes irrelevant. Considering that a data set consists of several classes as (C1, C2 ,...,Cn) and T represents the class values, the probability of a class is Pi= (Ci / |T|) and the entropy of the classes is;

$$\mathrm{Entropy(T)} = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{1}$$

Consider that the data set is divided into subsets as T class values (T1, T2 ,...,Tn) according to the B attribute. The gain to be obtained as a result of dividing T class values by using B attribute values;

$$\mathrm{Gain(B,T)} = Entropy(T) - \sum_{i=1}^{n} \frac{|T_i|}{|T|} Entropy(T_i) \tag{2}$$

The partitioning information is used to determine the value of attribute B for set T.

$$\text{Partitioning information(B)} = -\sum_{i=1}^{k} \frac{|T_i|}{|T|} log_2(\frac{|T_i|}{|T|}) \tag{3}$$

The gain rate is calculated as Eq 4.

$$\text{Gain rate} = \frac{\mathrm{Gain(B,T)}}{\text{Partitioning information(B)}} \tag{4}$$

The decision tree classifier has two phases: the build phase and the pruning phase (Breiman et al., 1984). In the first stage, build phase, a tree is created by dividing the training data set according to the best criterion until the data in each section has the same class label. Data over fitting may occur at this stage (Abdulsalam et al., 2015). The pruning phase provides a smaller tree size and thus improves the generalization accuracy rate (Kamber & Hand, 2006).

1.2.2. Decision Trees Algorithms

Decision tree algorithms have successful applications in many different fields. The data set obtained for personnel data has been subjected to analysis using random forest, J48, Logistic Model Tree, REPTree, Random tree, Decision Stump and Hoeffding tree. Among these algorithms, The J48 is one of the most well-known and widely used decision tree-based algorithms. This algorithm uses information to gain the ratio as the test attribute selection criteria. The information gain of each column in the dataset is divided by the information gain of the class to calculate the gain of each column. For each data set, the feature with the highest information gain rate is selected (Quinlan, 1993). The random tree algorithm has no pruning and creates a tree for a specific number of randomly selected properties at each node. The generated tree is randomly selected from the possible tree set. Each tree set has an equal chance of being tried (Fan et al., 2003). The random forest developed by Breiman is a supervised learning algorithm and performs well at classifying large amounts of data. This algorithm builds the decision tree by randomizing the divide at each node. It contains many single and unpruned decision trees. The class with the maximum value in the decision forest is picked as the final choice (Duggal, 2020). The Hoeffding tree method is a decision tree classification method that may be used successfully on enormous data sets by analysing each data set just once. For the building and evaluation of the decision tree, the method employs the Hoeffding bound (Domingos, 2000). REPTree algorithm is a decision tree classification method that is one of the fast. The algorithm includes the information gain criterion in building the decision tree (Zhao & Zhang, 2008). The Decision Stump algorithm is a decision tree classification method that creates a single level. In the tree created with this algorithm, the root node is directly connected to the leaf nodes (Witten & Frank, 2005). LMT (Logistic Model Trees) algorithm is a decision tree classification method that combines decision tree and logistic regression models. While ordinary decision trees form a piecewise fixed model with leaves, the logistic model tree is a decision tree with a linear regression model whose leaves provide a piecewise linear regression (Landwehr et al., 2005).

1.3. Performance Measures

The classification performance of classifiers on a data set is tested through cross-validation. In the present study, 10 fold cross validation was used. With 10 fold cross validation, the data is divided into training and testing data sets. In this present study, the data set is divided into 10 parts. Then one part is the test set and the remaining nine parts are the training set. After the classification process is done, the remaining ten pieces become the test set. The accuracy rate of each piece is calculated, and the classification success is calculated (Ghosh et al., 2015).

There are many different performance evaluation criteria to evaluate the success of decision tree algorithms. The complexity matrix is a summary of the estimated results of a classification process. It provides a summary of correct and incorrect predictions for each class. Four different outputs are obtained as a result of testing the classification model. These outputs (Tan et al., 2006); True Positive (TP): The value predicted as positive is actually positive. True Negative (TN): The negative predicted value is actually negative. False Positive (FP): A negative predicted value is actually positive. False Negative (FN): When the value predicted as positive is actually negative.

In this present study, accuracy, precision, sensitivity, F measure, kappa statistics and area under the ROC curve were used to evaluate the performance of decision tree algorithms. Accuracy, precision, sensitivity and F-measure formulas in Equations (5)-(8) is given.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FN+FP}} = \frac{\text{number of correctly classified samples}}{\text{total number of samples}} \tag{5}$$

The accuracy of an algorithm is calculated as the percentage of the data set correctly predicted by the algorithm.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+ FP}} = \frac{\text{number of correctly classified positive samples}}{\text{number of positively classified samples}} \tag{6}$$

The precision value indicates the number of samples classified as positive.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+ FP}} = \frac{\text{number of correctly classified positive samples}}{\text{number of positively classified samples}} \tag{7}$$

The recall value shows the rate at which a system detects positives.

$$\text{F} - \text{measure} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall+Precision}} \tag{8}$$

The F-measure could be considered as the harmonic precision mean.

The area under the ROC curve is obtained by plotting the change of the TP ratio compared to the FP ratio. The excess area under the curve indicates the success of the diagnostic test. Kappa coefficient is a statistical method that measures the agreement between two observers in evaluating categorical items (Cohen, 1960).

In addition, in this present study, mean absolute error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) are selected as the evaluation index to determine the error rates of classification algorithms. This formulas in Equations (9)-(12) is given.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_i - \hat{y_i}| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_i - \hat{y_i})^{2}} \tag{10}$$

$$RAE = \frac{\sum_{j=1}^{n}|y_i-\hat{y_i}|}{\sum_{j=1}^{n}|y_i-\bar{y}\,|} \tag{11}$$

$$RRSE = \sqrt{\frac{\sum_{j=1}^{n}(y_i-\hat{y_i})^{2}}{\sum_{j=1}^{n}(y_i-\bar{y}\,)^{2}}} \tag{12}$$

2. Finding and Discussions

This present study was carried out using the data set created based on personnel data received from the HR department. The reasons for leaving the job in the data set were analyzed according to the attributes of personnel collar type, age, educational status, and working hours. The performance metric values were obtained with random forest, J48, Logistic Model Trees (LMT), REPTree, random tree, Decision Stump, Hoeffding tree classification algorithms. Using a javabased workbench called Weka version 3.9.5, experiments are conducted on a Windows 11 machine with a Core (TM) i7 processor assessed by Nvidia Geforce RTX 3060 GPU and 16 GB of RAM. The performance metric values obtained are shown in Table 1. A 10-fold cross validation test was applied during the classification process.

Table 1

The Obtained Results of Classification Algorithms

	Accuracy	Kappa statistic	MAE	RMSE	RAE	RRSE
Random forest	91,42%	0,879	0,034	0,126	14,17%	36,53%
J48	90,02%	0,858	0,048	0,154	19,94%	44,67%
Logistic Model Tree	89,22%	0,847	0,055	0,161	22,95%	46,57%
REPTree	88,02%	0,829	0,053	0,162	22,10%	47,01%
Random tree	88,02%	0,877	0,053	0,162	22,09%	47,01%
Decision Stump	61,59%	0,381	0,192	0,310	80,49%	89,99%
Hoeffding tree	49,20%	0,271	0,212	0,330	89,11%	95,58%

Table 1 shows that the random tree algorithm achieves the best results (91.42%) in terms of accuracy. For real-world issues, the achieved accuracy is fairly high. The J48 algorithm provides a result of 90% and it is followed by LMT, Rep tree, random tree and decision stump respectively. The Hoeffding tree algorithm provides a result of 49.20%. The Hoeffding tree algorithm gave the lowest result.

Table 2 shows that the results of TP rate, FP rate, precision, recall, F-measure, MCC, ROC area performance metrics obtained for each class of each classification algorithm.

Table 2

The *Obtained Performance Criteria Results of Classification Algorithms for Each Class*

Randomforest							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0,967	0,051	0,931	0,967	0,948	0,996	Resignation
	0,883	0,037	0,908	0,883	0,895	0,991	Termination of employment
	1,00	0,00	1,00	1,00	1,00	1,00	Pension
	0,771	0,026	0,831	0,771	0,8	0,987	Discontinuity
	0,97	0,004	0,941	0,97	0,955	1,00	Military/Marriage
	0,895	0,004	0,90	0,895	0,895	0,999	Abuse of trust
Weighted Avg	0,914	0,036	0,913	0,914	0,913	0,994	
J48							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0,952	0,058	0,921	0,952	0,936	0,992	Resignation
	0,952	0,087	0,817	0,952	0,879	0,972	Termination of employment
	1,00	0,00	1,00	1,00	1,00	1,00	Pension
	0,614	0,002	0,977	0,614	0,754	0,959	Discontinuity
	0,939	0,002	0,969	0,939	0,954	1,00	Military/Marriage
	0,789	0,00	1,00	0,789	0,882	0,998	Abuse of trust
Weighted Avg	0,9	0,05	0,909	0,9	0,897	0,983	
LMT							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0,928	0,051	0,928	0,928	0,928	0,99	Resignation
	0,959	0,093	0,808	0,959	0,877	0,977	Termination of employment
	1,00	0,00	1,00	1,00	1,00	1,00	Pension
	0,629	0,002	0,978	0,629	0,765	0,97	Discontinuity
	0,909	0,009	0,882	0,909	0,896	1,00	Military/Marriage
	0,789	0,002	0,94	0,789	0,857	0,997	Abuse of trust
Weighted Avg	0,892	0,049	0,901	0,892	0,889	0,985	
Reptree							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0,962	0,099	0,874	0,962	0,916	0,986	Resignation
	0,903	0,067	0,845	0,903	0,873	0,98	Termination of employment

	1,00	0,00	1,00	1,00	1,00	1,00	Pension
	0,614	0,009	0,915	0,614	0,735	0,973	Discontinuity
	0,848	0,004	0,933	0,848	0,889	0,997	Military/Marriage
	0,684	0,002	0,929	0,684	0,788	0,995	Abuse of trust
Weighted Avg	0,88	0,063	0,884	0,88	0,876	0,984	
Randomtree							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0,99	0,072	0,908	0,99	0,947	0,996	Resignation
	0,924	0,056	0,87	0,924	0,896	0,991	Termination of employment
	1,00	0,00	1,00	1,00	1,00	1,00	Pension
	0,657	0,002	0,979	0,657	0,786	0,987	Discontinuity
	0,939	0,002	0,969	0,939	0,954	1,00	Military/Marriage
	0,789	0,00	1,00	0,789	0,882	0,999	Abuse of trust
Weighted Avg	0,914	0,047	0,919	0,914	0,911	0,994	
Decisionstump							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0,934	0,511	0,564	0,934	0,703	0,712	Resignation
	0,754	0,122	0,73	0,754	0,742	0,816	Termination of employment
	0,00	0,00	-	0,00	-	0,67	Pension
	0,00	0,00	-	0,00	-	0,54	Discontinuity
	0,00	0,00	-	0,00	-	0,667	Military/Marriage
	0,00	0,00	-	0,00	-	0,577	Abuse of trust
Weighted Avg	0,616	0,248	-	0,616	-	0,71	
Hoeffding tree							
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	0,641	0,376	0,55	0,641	0,592	0,653	Resignation
	0,459	0,136	0,58	0,459	0,512	0,774	Termination of employment
	0,82	0,03	0,61	0,82	0,70	0,98	Pension
	0,361	0,189	0,237	0,361	0,286	0,689	Discontinuity
	0,008	0,004	0,111	0,008	0,014	0,88	Military/Marriage
	0,00	0,00	-	0,00	-	0,772	Abuse of trust
Weighted Avg	0,492	0,224	-	0,492	-	0,729	

The extracted best rules by the chosed algorithm are illustrated in Table 3. The obtained rules were obtained using the Random forest algorithm, which gives the highest accuracy rate of 91.42%.

Table 3

Examples of Extracted Rules

Rule 1	IF (Education= Primary school) and (42<Age<44 and (3.5<= Operation time (months) < 4.5), THEN class Abuse of trust.
Rule 2	IF (Education= Primary school) and (42<Age<45) and (4.5<= Operation time (months) <6), THEN class is Discontinuity.
Rule 3	IF (Education= Primary school) and (6< Operation time (months) <7.5) and (18<Age<19), THEN class Military service/marriage.
Rule 4	IF (Education= High school) and (Operation time (months) < 1.5) and (18<Age<24), THEN class is Termination of employment.
Rule 5	IF (Education= High school) and (Operation time (months) < 2.5) and (33<Age< 36), THEN class Resignation.
Rule 6	IF (Education= High school) and (Operation time (months) < 1.5) and (30<Age<33), THEN class is Discontinuity.

Rule 7	IF (Education= High school) and (1.5< Operation time (months) <2.5) and (18<Age<24), THEN class is Resignation.
Rule 8	IF (Education=Junior college) and (11.5< Operation time (months) < 22.5) and (Age < 24), THEN class is Resignation.
Rule 9	IF (Education= Bachelor/Master) and (Operation time (months) < 2.5) and (Age < 36), THEN class is Termination of employment.
Rule 10	IF (Education= Bachelor/Master) and (17.5< Operation time (months) < 19.5) and (23<Age<24), THEN class is Resignation.
Rule 11	IF (Education= Junior college) and (36<Age<39) and (Operation time (months) < 4), THEN class is Termination of employment.
Rule 12	IF (Education= Primary school) and (Age>48) and (32.5<= Operation time (months) <33.5), THEN class is Retirement.
Rule 13	IF (Education=Primary school) and (1.5< Operation time (months) <11.5) and (48<Age<51), THEN class is Discontinuity.
Rule 14	IF (Education=Bachelor/Master) and (Operation time (months)>=33.5) and (Age>38), THEN class Retirement.
Rule 15	IF (Education=Primary school) and (19.5< Operation time (months) < 25) and (Age< 23), THEN class Military service/marriage.
Rule 16	IF (Education= High school) and (Operation time (months) < 13.5) and (25<=Age< 31), THEN class Abuse of trust.
Rule 17	IF (Education= Junior college) and (Operation time (months) < 6) and (Age< 25), THEN class is Discontinuity.

When the rules are examined in general, it has been observed that the reasons for leaving the job for individuals whose education level is primary school and age 42-45 are discontinuity or abuse of trust. If the education level is bachelor/master and the operation year is less than 1 year, the reason for leaving the job is resignation. If the working period is more than 30 months and the age is over 38, the reason for leaving the job is retirement.

According the rules, for example in rule 1, if the education level is primary school, age is 42-44, and operation time is 3.5-4.5, the reason for leaving the job is abuse of trust. In rule 2, if the education level is primary school, age is 42-45, and operation time is 4.5-6, the reason for leaving the job is discontinuity. In rule 3, if the education level is primary school, the age is 18-19, and the operation time is less than 7.5 months, the reason for leaving the job is service or marriage. In rules 5 and 7, if the education level is high school, the age is 18-24 or 33-36, and the operation time is less than 2.5 months, the reason for leaving the job is resignation. In rule 9, if the education level is bachelor/master, the age is less than 36, and the operation time is less than 2.5 months, the reason for leaving the job is termination of employment. In rule 17, if the education level is junior college, the age is less than 25, and the operation time is less than 6 months, the reason for leaving the job is discontinuity.

Conclusion, Discussion and Recommendations

Technological developments have played an important role in the development and change of the world. With digitalization in human resources management, it has reduced the burden on the administrative staff in the company, minimized biased behavior in decision-making, and enabled the prediction of employee retention rates. In addition, the development of computers has brought data exchange to significant levels. As a result, the production of more and more dense data has revealed the concept of big data and made algorithm software important for processing this data.

The purpose of this study and its contribution to the literature is to examine the impact of digitalization on human resources management practices and to determine trends in this regard. In this context, the reasons for dismissal were examined, classification algorithms in the literature were compared, and the best rule-based algorithm was selected by classifying the reasons for dismissal. When the results obtained were examined, the accuracy values obtained from the Random forest algorithm gave the best result with 91.42%. Accuracy values are respectively; J48 algorithm was 90.02%, Logistic Model Tree algorithm was 89.22%, REPTree algorithm was 88.02%, Random tree algorithm was 88.02%, Decision Stump algorithm was 61.59%. Hoeffding tree algorithm gave the lowest result with 49.20%

In this study, in addition, decision trees-based classification algorithms were used. In future studies, statistical-based classification (such as Bayesian, regression) or distance-based classification algorithms (such as nearest neighbor) can be used and can also be applied in performance evaluation and recruitment processes.

Statement of Research and Publication Ethics	The article does not require ethics committee approval.
Conflict of Interest	There is no potential conflict of interest in the study.
Acknowledgment or Statement of Support	No support is taken from any institution or organization.
Author Contributions	The first author contributes 35% and the second author contributes 35% while the third author contributes 30%.

Kaynakça

Abdulsalam, S. O., Babatunde, A. N., Hambali, M. A., & Babatunde, R. S. (2015). Comparative analysis of decision tree algorithms for predicting undergraduate students' performance in computer programming. *Journal of Advances in Scientific Research & Its Application* (JASRA), *2*, 79-92.

Akduman, G. (2019). Digital recruitment: Evaluation of the impact of the digital world on the human resources recruitment function with conceptual and application examples. *International Journal of Arts and Social Studies, 2*(3), 24-44.

Amershi, S., & Conati, C. (2006). Automatic recognition of learner groups in exploratory learning environments. In: Ikeda, M., Ashley, K.D., Chan, TW. (Eds.) Intelligent Tutoring Systems. Lecture Notes in Computer Science, 4053. Springer

Argüden, Y., & Erşahin, B. (2008). Veri madenciliği: Veriden bilgiye, masraftan değere. Alkim Kağıt Sanayi ve Ticaret A. Ş.

Bongiorno, G., Rizzo, D., & Vaia, G. (2018). CIOs and the digital transformation: A new leadership role (pp. 1-9). *Springer International Publishing.*

Breiman L., Friedman J. H., Olshen R. A. & Stone C. J. (1984). *Classification and regression trees*. Wadsworth.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46. https://doi.org/10.1177/001316446002000104.

Domingos, P., & Hulten, G. (2000). Mining high-speed data streams. *In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 71-80. https://doi.org/10.1145/347090.347107

Duggal, P., & Shukla, S. (2020). Prediction of thyroid disorders using advanced machine learning techniques. *In 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 670-675). IEEE. https://doi.org/0.1109/Confluence47617.2020.9058102

Erdem, S., & Özdağoğlu, G. (2008). Analyzing of emergency data of a training and research hospital in aegean region using data mining. *Anadolu University Journal of Science and Technology, 9*(2), 261-270.

Fan, W., Wang, H., Yu, P. S., & Ma, S. (2003). Is random model better? on its accuracy and efficiency. *In Third IEEE International Conference on Data Mining* (pp. 51-58). IEEE. https://doi.org/10.1109/ICDM.2003.1250902

Filizöz, B., & Orhan, U. (2018). Industry 4.0 in the context of human resources management: A literature study. *Cumhuriyet University Journal of Economics and Administrative Sciences, 19*(2), 110-117.

Ghosh, A., Guha, T., & Bhar, R. B. (2015). Identification of handloom and powerloom fabrics using proximal support vector machines. *Indian Journal of Fibre & Textile Research*, *40*(1), 87-93. https://doi.org/10.56042/ijftr.v40i1.3809

Han J., & Kamber M. (2006). *Data mining: Concepts and techniques*. Morgan Kaufmann, Elsevier Science.

Kalıkov, A. (2006). *Veri madenciliği ve bir e-ticaret uygulaması* [Yüksek Lisans Tezi]. Gazi Üniversitesi.

Karaboğa, T., & Karaboğa, H. A. (2022). Digitalization in human resources management: A bibliometric review. *Turkish Studies-Economics*, *Finance, Politics*, *17*(2), 343-364. https://doi.org/10.7827/TurkishStudies.58236

Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine learning*, *59*, 161-205.

Maimon, O. Z., & Rokach, L. (2005). *Data mining and knowledge discovery handbook*, Springer.

Mingers, J. (1989). An empirical comparison of pruning methods for decision tree induction. *Machine Learning, 4*, 227-243.

Mulyana, R., Rusu, L., & Perjons, E. (2021). It governance mechanisms influence on digital transformation: A systematic literature review. *In Twenty-Seventh Americas' Conference on Information Systems (AMCIS), Digital Innovation and Entrepreneurship, Virtual Conference, August 9-13*, 2021 (pp. 1-10). Association for Information Systems (AIS).

Quinlan, J. R. (1993). *Programs for machine learning*, 4(5).

Ruël, H.J.M., Bondarouk, T.V., & Van der Velde, M. (2007). The contribution of e-HRM to HRM effectiveness. *Employee Relations, 29*(3), 280-291. https://doi.org/10.1108/01425450710741757

Ruël, H.J.M., & Bondarouk, T. V. (2009). Electronic human resource management: Challenges in the digital era. *The International Journal of Human Resource Management, 20*(3), 505-514. https://doi.org/10.1080/09585190802707235

Shen, Y. (2007). *A formal ontology for data mining: principles, design, and evolution* [Doctoral dissertation], Université du Québec à Trois-Rivières.

Shen, Y., Xing, L., & Peng, Y. (2007). Study and application of Web-based data mining in e-business. In *eighth ACIS international conference on software engineering, artificial intelligence, networking, and parallel/distributed computing*, *1*, 812-816. IEEE. https://doi.org/10.1109/SNPD.2007.117

Synnestvedt, M. B., Chen, C., & Holmes, J. H. (2005). CiteSpace II: Visualization and knowledge discovery in bibliographic databases. In *AMIA annual symposium proceedings* (2005, p. 724). American Medical Informatics Association.

Tan, K. C., Yu, Q., & Ang, J. H. (2006). A dual-objective evolutionary algorithm for rules extraction in data mining. *Computational optimization and applications,* 34, 273-294. https://doi.org/10.1007/s10589-005-3907-9

Uğurlu, H. Ü. A., & Doğan, A. (2023). Digital transformation and digitalized recruitment function in human resources management. *Kocaeli University Journal of Social Sciences, 1*(45), 1-16. https://doi.org/10.35343/kosbed.1247587

Ulrich, D., & Dulebohn, J. H. (2015). Are we there yet? What's next for HR?. *Human Resource Management Review, 25*(2), 188-204. https://doi.org/10.1016/j.hrmr.2015.01.004

Van Grembergen, W., & De Haes, S. (2009). *Enterprise governance of information technology: Achieving strategic alignment and value*. Springer Publishing Company, Incorporated.

Vial, G. (2021). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118-144. https://doi.org/10.1016/j.jsis.2019.01.003

Witten, I. H., & Frank, E. (2002). Data mining: Practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record, 31*(1), 76-77. https://doi.org/10.1145/507338.507355

Yıldırım, P., Birant, D., & Alpyildiz, T. (2018). Data mining and machine learning in textile industry. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(1), e1228. https://doi.org/10.1002/widm.1228

Yılmaz, C., & Yılmaz, T. (2023). The effect of industry 4.0 on human resources management: HRM 4.0. *Hak İş International Journal of Labor and Society, 12*(32), 11-28. https://doi.org/10.31199/hakisderg.1214130

Zhao, Y., & Zhang, Y. (2008). Comparison of decision tree methods for finding active objects. *Advances in Space Research, 41*(12), 1955-1959. https://doi.org/10.1016/j.asr.2007.07.020