



1 Introduction

Human resource analytics is a special part of analytics where the main focus is the human resource. In HR analytics, the analytical process is applied to the organization's human resources. The main objective of this process is not only to enhance employee performance but also to improve overall employee satisfaction. HR analytics helps institutions by providing useful insights for the human resource functions by collecting data and transforming it into useful information for empowering human resource-related processes. HR analytics is very useful for improving HR functions such as workforce management, recruitment, performance evaluation, and development management. In this chapter, four different HR problems are focused, and these problems are tried to solve by using HR analytics tools. The rest of the chapter is organized as follows. Chapter two gives the four different HR problems, namely attrition risk, recruitment, performance evaluation, and training planning. These problems are defined via a comprehensive literature review. For each problem, a case study is defined and HR analytics tools are applied. The last section concludes and gives further suggestions.

O. Akarsu

Koc Digital, Camlica Business Center No: 11, 34700 Unalan, Uskudar, Istanbul, Turkey

C. Kadaifci · S. C. Onar (✉)

Department of Industrial Engineering, Faculty of Management, Istanbul Technical University, Macka, 34367 Istanbul, Turkey

e-mail: cevikse@itu.edu.tr

2 Employee Turnover (Attrition)

Employee turnover (i.e., attrition) refers to the number (or ratio) of employees leaving the organization either voluntarily or involuntarily. Voluntary turnover is based on the decision of employees, while the involuntary one occurs when an employer decides to end the employment relationship, as a result of the death or retirement of the employee [1, 2]. From an organization's perspective, performance under expectations, improper behavior, or adaptation problems regarding the culture or working environment may cause the termination of the employment agreement [1, 3]. Since the control of the organizations over the employee-related involuntary turnover is low, in the scope of this study, the voluntary turnover will be examined.

Starting from recruitment, organizations invest in each employee through selection, orientation, training, improvement, and positioning. When an employee leaves the organization, then the position should be filled immediately with an appropriate candidate. Accordingly, an employee loss causes a significant cost as well as affecting the organizational value [4], the long-term strategies [5], the competitive advantage [1, 6], the organizational performance [7, 8], and even loss of customers, especially if the leaving employee is a talented and highly capable one. Being an important factor affecting the organization's overall performance, employee turnover deserves considerable attention from both of the organizations and the researchers in order to enable taking precautions to diminish (or manage) the employee loss. Thanks to the developments and improvements in the learning-based data analysis techniques, the employee data representing the reasons for the voluntary employee attrition can be examined, and whether the employee quits or not can be predicted. The common approach in the recent literature on the prediction of employee turnover is applying different methods or algorithms and comparing them based on the performance measures such as accuracy, precision, recall, specificity, the area under the curve (AUC), and F-measure. In Table 1, some studies are provided along with the performance measures and the dataset utilized.

According to a detailed literature review, gender [1, 6, 8–10], state of origin [1], duration of service [1, 4, 8–10], the title in the organization or job level [1, 6, 8, 9], annual or monthly salary [1, 4, 8, 9], job satisfaction [4, 6, 8–10], environment satisfaction [6, 8–10], relationship satisfaction [8, 9], performance of the employee [4, 6, 8, 9], work–life balance [6, 8–10], the number of trainings the employee participated in the previous year [8, 9], promotion [4], feeling appreciated [10] are among the factors considered for the prediction. Even though several algorithms and methods including support vector machines [4, 5, 8, 9, 11–13], AdaBoost [4, 8], neural networks [13], KNN [8, 9, 11, 12], and naïve Bayes [5, 8, 9, 11, 12] are used, the methods preferred considering the performance measures are decision tree [1, 4, 8, 9, 11–13], random forest [4, 5, 8–10, 12, 13], and XGBoost [5, 11, 13, 14].

To achieve a well-managed prevention process to diminish employee attrition, the organizations better consider the most important factors affecting employees' quitting decisions. According to recent studies, the relatively high attrition is mostly related to the lower involvement [10], motivation, and satisfaction of employees [9];

Table 1 Learning-based methods to predict employee attrition

Methods	Accuracy	Precision	Recall/Sensitivity	F-measure	AUC	Specificity	Dataset	N	References
Decision tree (C4.5 (J48) classifier)	0.670	0.613	0.670	0.636	0.784	-	Higher Institutions in Nigeria	309	[1]
Logistic regression	0.871	-	0.871	0.853	0.816	0.454	IBM HR analytics	-	[8]
Decision tree	0.982	0.950	0.975	0.962	-	-	-	14,999	[4]
Random forest (with feature selection)	0.986	0.993	0.948	0.970	-	-	-	-	[4]
Logistic regression	0.750	0.356	0.733	-	0.743	0.753	-	4410	[6]
Random forest	0.808	-	-	-	0.851	-	IBM HR analytics	1470	[9]
Voting classifier (with feature selection)	0.960	-	-	0.620	-	-	IBM HR analytics	1470	[13]
Voting classifier	0.980	-	-	0.880	-	-	Kaggle HR dataset	15,000	[13]
Voting classifier (with feature selection)	0.990	-	-	0.910	-	-	Small-sized HR real data	450	[13]
Random forest	0.988	0.983	0.994	0.990	-	-	Real data	1649	[10]
XGBoost	0.960	-	0.950	0.960	0.950	-	-	-	[11]
XGBoost	0.891	-	-	-	-	-	IBM HR analytics	1470	[14]
XGBoost	0.880	-	-	-	-	-	HR Information Systems (HRIS) database and Bureau of Labor Statistics	73,115	[5]
Random forest	0.990	0.990	0.989	0.993	-	0.994	IBM HR Analytics	1470	[12]

thus, improving the working environment, balancing the workload, and sustaining a strong relationship between the employee and management may help to deal with this problem [6, 10]. Using existing employee data to extract invaluable information may support organizations to enhance their decision-making capabilities regarding employment strategies.

Case Study: Employee Turnover (Attrition)

In this use case, a machine learning model was built to predict employee attrition by using an employee dataset. The dataset is a modified version of IBM HR Analytics Employee Attrition Data [50]. The main research question is to figure what is the attrition risk of my current employees and what are the factors that affect employee attrition?

The dataset includes 20 variables of 1102 employees in a company. The list of variables and their definitions can be seen in Table 2.

For the next step, unnecessary variables were removed, and a brief exploratory data analysis was conducted. Since the main goal of this analytics model was to predict

Table 2 List of variables_employee attrition

Variable name	Variable description
EmpID	Unique ID of employees
Age	Age of the employee
Daily rate	The daily wage rate of an employee
Distance from home	The distance between the household and company location
Education	Education level of the employee (1: High School,3: Undergraduate, 4: Masters, 5: Ph.D.)
MonthlyIncome	The monthly wage of an employee
NumCompaniesWorked	Number of companies that employee worked before
PerformanceRating	Last performance score of the employee
RelationshipSatisfaction	Satisfaction score of the employee
TrainingTimesLastYear	Number of trainings employee took during last year
YearsInCurrentRole	How many employee spent in the current role
YearsSinceLastPromotion	How many years have passed since the last promotion
YearsWithCurrManager	How many years passed with the current manager
Tenure	Total number of years within the company
TotalExperience	Total number of years in the work-life
WageRate	Comparison of the wage with the company mean
WorkingModel	Working model of the employee (Hybrid, working in the office, working remote)
OrgFunction	Organizational department of the employee
PositionEngagement	Engagement score for the position
Attrition	Active or former employee (0: Active, 1: Former employee)

the attrition risk of employees and identify the reasons behind attrition, a bivariate analysis was proceeded to understand the distribution of attrition. To conduct the bivariate analysis, the data types of the features were modified and checked. The dataset included 4 categorical, 16 numeric, and 1 string features. Since the feature “EmpID” represents the unique ID of each employee, it was removed from the analysis. Additionally, null values were checked for each column. It was seen that the dataset did not include any null value. The attrition ratios of former and current employees for each categorical variable is compared by using bi-variate analysis. The results of this analysis can be given as follows:

- For the education level = 1, attrition level %17.3 is slightly higher than the overall ratio (%15.6)
- For the education level = 5, attrition level %7.4 is lower than the overall ratio (%15.60)
- The employees whose wages are less than the company mean (Wage rate = 1) have a slightly higher attrition ratio (%17.9) than the other ones
- The employees who are working remote (Working Model = 2) have a higher attrition rate (%25.4) than the other groups
- Employees working in the office (Working Model = 1) have the lowest attrition level when compared with the other two working models.
- Employees from departments “0” and “2” have slightly higher attrition ratios than the expected mean.

As a result of the exploratory data analysis, it is possible to say that all of the features within the dataset could be used for modeling. Possible signs of relations between target column “Attrition” and dependent variables such as “Age,” “Monthly Income,” “Years in current role,” “Education Level,” and “Working Model” are possible predictors of attrition of employees. However, since these comparisons were conducted on variable levels separately, the outputs of the model should be analyzed to understand the complex relationships.

For the next step, a classification model was applied to predict the attrition risk of current employees. All of the variables except “EmpID” were used as predictors and the variable “Attrition” was used as the target. In this study, three different machine learning algorithms were applied to the Attrition dataset. The results showed that “LightGBM” classifier outperformed the other two algorithms. Since data is imbalanced with an attrition rate of 15.6% checking, both accuracy and F1 scores are crucial to assess model performance. Even though the decision tree model is a simpler model than random forest and LightGBM, it outperformed the other two with a recall score 0.34 and F1 score 0.32. However, its overall accuracy is worse than the other two (0.76). Random forest model has a higher accuracy rate, but the recall metric is considerably low (0.09). Nonetheless, F1 Score is 0.16. LightGBM has a better F1 score (0.27) than random forest. Additionally, the recall metric for LightGBM is also higher than random forest (0.18). Additionally, in terms of overall accuracy, it outperformed the other two with a score of 0.84. As a result, the LightGBM algorithm was chosen to predict the employee attrition risk with the best performance metrics.

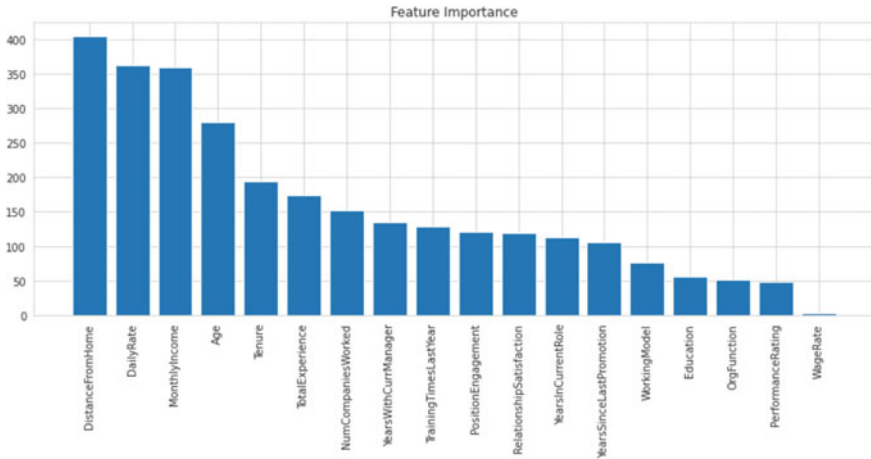


Fig. 1 Future importance_employee attrition

For the next step, the feature importance of the LightGBM model was calculated. The feature importance plot is given in Fig. 1.

Figure 1 shows that “Distance from Home,” “DailyRate,” and “MonthlyIncome” are the most important three features for the classification model.

Deep dive analysis provided that if the Dailyrate decreases, the attrition risk of employees increases. Lower monthly income increases the risk of attrition. Low levels of distance from home decrease the risk of attrition.

3 Recruitment Analytics

Organizations aim to hire appropriate employees having the appropriate knowledge, skills, and abilities to contribute to the success and sustain the competitive advantage, ensuring the right fit for the vacant positions is in the scope of the human resources recruitment function.

Recruitment, consisting of activities to be performed to find and attract qualified employees, affects the organizational performance through the success of the subsequent HR functions such as selection, compensation, and training [15]. Recruiting requires considerable knowledge and expertise, especially as it is a data-driven and knowledge-sensitive process, to attract the job seekers to apply to the job openings, maintain the interest of job seekers in the position as long as being open, review the curriculum vitae (CV) of the candidates, and match the requirements of the com company with potential applicants [16, 17]. As involving a huge amount of

data, automatizing the process is among the primary concerns of both the practitioners and researchers. Turning the existing data into invaluable business intelligence, recommender systems become appealing both to the recruiters and to the job seekers.

In order to offer job seekers positions attracting their attention, job recommender systems are developed. For instance, to match job seekers with available job openings, a career platform classifies both the CVs of the registered users and available job titles using support vector machines and k-nearest neighbor methods [18]. A job recommendation system consisting of random forest and support vector machines is proposed to target the price and location sensitive people separately [19].

To support recruiters, an automated recruitment system imitating the knowledge of the recruiter, extracting data from CVs, and matching the requirements and profiles is built [17]. In a study proposing a new data-based design for electronic human resources management (e-HRM) activities including e-recruitment, e-selection, e-performance management, e-compensation, and e-learning, attrition is used to predict potential positions to be input to the e-recruitment system using decision tree [classification and regression trees (CART)], support vector machines, and k-nearest neighbor methods [20]. Decision tree performs better compared to other methods. The effectiveness and usability of these systems are quite important for their prevalence. Accordingly, information quality, popularity, and security of the source are found to be the factors affecting the effectiveness of e-recruitment [21].

Recruitment sources should be examined carefully due to their effect on employee performance, employee turnover [22], or employee loyalty [23]. Even though there is not any consensus on the sources resulting in the recruitment of employees performing better, several studies are conducted to understand the nature of the relationship. The primary reasons of differences in the prominent sources of recruitment can be summarized as follows: (i) using a different number of sources [24] and even combining them into different groups (e.g., internal and external [25], being controllable or uncontrollable by the organization [16]), (ii) assuming that the job seekers and/or recruiters use single source and neglecting the usage of multiple sources [24], (iii) the differences in the employee demographics or characteristics, (iv) the differences in the type of jobs (i.e., white/blue-collar positions, permanent/temporary jobs, etc.) [24], (v) the differences in the type of organization and industry (i.e., local/global [26]), (vi) the differences in sample sizes [24].

An organization might fill existing or expected vacancies internally by recruiting its own employees or externally [16]. Interdepartmental transfers, promotions, and internship programs are considered in internal sources while external sources of recruiting include recommendations, employment agencies, universities and colleges, referral programs, job fairs, professional associations, and trade unions, rehiring former employees, and advertisements, as well as employee exchange [24–26].

Performance-related studies indicate that internal sources have a significant effect on job performance [25], employees hired using employee referral have higher job performance than those hired through other methods [27], and agencies do better in finding the appropriate employees [28].

Table 3 Sample dataset recruitment

	EmpID	PerformanceScore	OrgEngagement	PositionEngagement	HireSource
0	Emp1	18.5	0.87	0.02	Applied online
1	Emp2	35.3	0.18	0.39	Campus
2	Emp3	67.9	2.75	3.11	Referral
3	Emp4	35	1.7	0.92	Campus
4	Emp5	27.5	0.35	0.38	Referral

Case Study: Assessing Recruitment Source Through Employee Performance and Engagement

In this case, an analytics approach was applied to an HR dataset to assess the best medium for hiring. The dataset used in this case is a modified version of IBM HR Analytics Employee Attrition Data [50]. The main objective is to assess recruitment sources through employee performance and engagement scores and define which recruitment source is better to ensure employee performance.

The dataset includes Performance Score, Organizational Engagement Score, Positional Engagement Score, and source of hiring for 852 employees in a company. A sample dataset is given in Table 3.

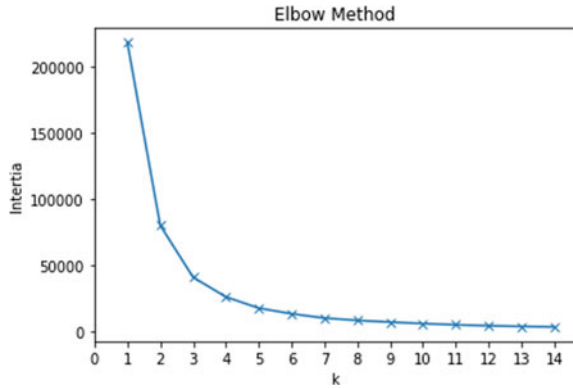
Brief statistical analysis of numerical variables showed that the performance score deviates between 0 and 100, whereas engagement scores deviate between 0 and 5. If the correlation between numerical features is checked, it is seen that Organizational and Positional Engagement scores are highly correlated (0.83). Additionally, the Performance score is correlated with Position Engagement with a correlation coefficient 0.65 but less correlated with Organizational Engagement with a coefficient of 0.54.

To select the best hiring source, the target variable using engagement and performance scores is defined. Since organizational engagement is related to employee commitment to the company and positional engagement is related to employee's willingness to stay in the current role, both of them are linked with long-term individual performance. By using these three indicators, employee data is segmented and every employee is classified in a segment that represents their engagement and performance scores. The segmentation is conducted by applying an unsupervised machine learning algorithm and on the next step, hiring sources are evaluated according to these created segments.

By using K-means clustering algorithm and elbow method, the optimum number of employee segments is determined as 4 (Fig. 2).

After the cluster analysis the employee groups are defined as follows: employees with a low level of engagement and performance as "Low," employees with mid-level performance level and mid-level engagement as "Medium," and employees with high performance and high engagement as "High". Since engagement is an important factor for sustainable performance, it is possible to assert that hiring source that provides more employees in class "High" is a better method for recruiting engaged

Fig. 2 K-means analysis-recruitment analysis



and high-performer employees. The distribution of hiring sources and the employee clusters are given in Table 4.

When the total distribution of high performers are compared for each hiring source, it is possible to say that “Search Firm” has the highest high performer ratio with %34.4 which is higher than the overall average (%27.5). On the other hand, it is possible to say that the hiring source “Referrals” has the lowest high performer ratio with %22.5 which is lower than the overall average (%27.5). As a result, a clustering approach based on the performance and engagement scores of employees showed us that “Search Firm” is the best hiring source to ensure long-term success within the company.

Table 4 The distribution of hire source and the clusters

HireSource	Perf_Cluster	
Applied online	High	29.1
	Low	26.5
	Mid	44.4
Campus	High	27.6
	Low	30.0
	Mid	42.4
Referral	High	22.5
	Low	26.4
	Mid	51.1
Search Firm	High	34.4
	Low	21.9
	Mid	43.8

4 Performance Analytics

Performance management, an approach for enhancing the overall targets of the organization by improving the performance of the employees (both individuals and teams) to benefit from their potential, ensures that the employees embrace the values and the goals of the organization [29]. An effective performance management approach requires measuring employee performance using appropriate tools and techniques to provide input to other HR functions such as career planning and training, develop strategies, or take necessary actions related to possible risks.

Employee performance directly affects the success, survival, and competitive power of the organizations. So, it is expected that the performance assessments support the managerial decisions by providing the necessary information. Considering the dynamic and rapidly changing nature of the business environment, the costly and time-consuming traditional performance assessment tools and techniques may fail to provide instant real-time information to support the decisions [30, 31]. At this point, data analytics methods and algorithms are used to observe the actual performance and even take precautions for future deteriorations likely to be based on continuous predictions [32].

In order to predict the employee performance, stochastic gradient descent and bagging classifier [33], neural networks [34], decision tree [33, 35–37], naïve Bayes [34, 36, 38], logistic regression [33, 34, 36, 39], random forest [33, 34, 39], and support vector machines [34, 39] are used. As given in Table 5, classification methods and algorithms are preferred. As the employee characteristics and the nature of the work have an impact on the performance, these studies consider various factors including gender, age, education, marital status, industry, organizational support, total experience in years, title of the position (e.g., managerial or non-managerial), leadership type, nature of the task (e.g., complexity), socio-economic status, location, income group, wage system, workspace and environment, task complexity, motivation, and competence, [34–37]. Personality-related factors such as the abilities of creative thinking, conflict resolution, decision-making, and personal relationships are also included in the models [34].

With the rise of information technologies, digitalization, and other related technologies [31], performance management activities are evolving to become more artificial intelligence-oriented (AI). Using AI is expected to result in continuous, real time, flawless data with a robust and unbiased performance management process [40].

Case Study: Who are the best performers? Identifying top manager profile through analytics

In this case, an analytics approach was applied to assess the profile of top-performer managers using an employee dataset. The dataset is a modified version of IBM HR Analytics Employee Attrition Data [50]. The objective of this case is to identify the top performer managers and their common characteristics.

Table 5 Methods to predict employee performance

Methods	Accuracy	Precision	Recall/Sensitivity	F-measure	AUC	N	References
Support vector machines	0.853	0.870	0.850	–	0.890	1500	[34]
Decision tree (CART)	0.906	–	–	–	–	120	[37]
Logistic regression	0.756	–	–	–	–	227	[39]
Decision tree	0.708	0.727	0.708	0.708	0.859	206	[35]
Naïve Bayes	0.806	0.804	0.806	0.805	0.894	1037	[38]
Logistic regression	0.834	0.843	1.000	0.915	0.799	858	[36]
Logistic regression	0.690	–	–	–	–	512	[33]
K-means clustering and artificial neural network (ANN)	–	–	–	–	–	220	[41]

The dataset includes 8 variables of 218 managers in a company. The list of variables and their definitions can be seen in Table 6.

To identify top performer managers within the company, evaluation scores of subordinates and business performance outcomes were analyzed. Basic descriptive statistics show that TeamAssessment has a mean of 2.24, and PerformanceScore has

Table 6 List of variables and their definitions

Variable name	Variable Description
Age	Age of the manager
TotalExperience	Total number of years in the work life
Education	Education level of the manager (1: High School,3: Undergraduate, 4: Masters, 5: Ph.D.)
EducationField	Graduation major of the manager
OrgFunction	Organizational department of the manager
Tenure	Total years of employment within the company
TeamAssessment	Team evaluation score of the manager
PerformanceScore	Performance evaluation score of the manager

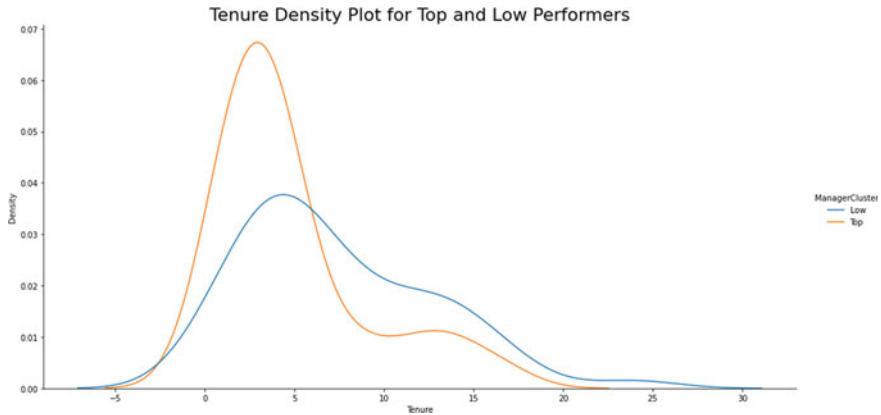


Fig. 3 The tenure density plot for top and low performers

54.2. The correlation coefficient between TeamAssessment and Performance Score is -0.0089 which indicates that these two variables are not correlated. To identify the top performer managers, firstly, managers who have higher performance scores than the company average were identified. Secondly, managers whose team assessment scores were higher than the company average were selected.

Analysis results proved that there are 60 top performers and 55 low-performer managers. Comparison of basic descriptive statistics of top and low performers show that top performers' average age is lower than low performers (41.3–37.3). Secondly, top performers have a higher total experience average and a lower tenure than low performers. Figure 3 illustrates the tenure density plot for top and low performers.

The density plot showed that the number of top performers who have a tenure less than 5 is higher than the low performers. On the contrary, the number of low performers with a tenure of more than 5 is higher than the top performers. Even though distributions of top and low performers in terms of total experience are similar, more top performers have a total experience of more than 15. When the top and low performers' education levels are compared, it is seen that the number of top performers who have a Master's degree is more than the low performers. On the contrary, the number of low performers who have a high school degree is higher than the top performers.

Analysis of the education field marks a considerable difference between technical degree and life sciences graduates. Figure 4 shows the performance difference between education degrees.

The number of top performers with a technical degree is much higher than the low performers. On the contrary, the number of low performers with life sciences degrees is much higher than the top performers. When the number of top and low performers are compared in different organizational departments, it is seen that the number of top performers is slightly higher in information technologies and lower in operations. In summary, the common characteristics of the top performers are they

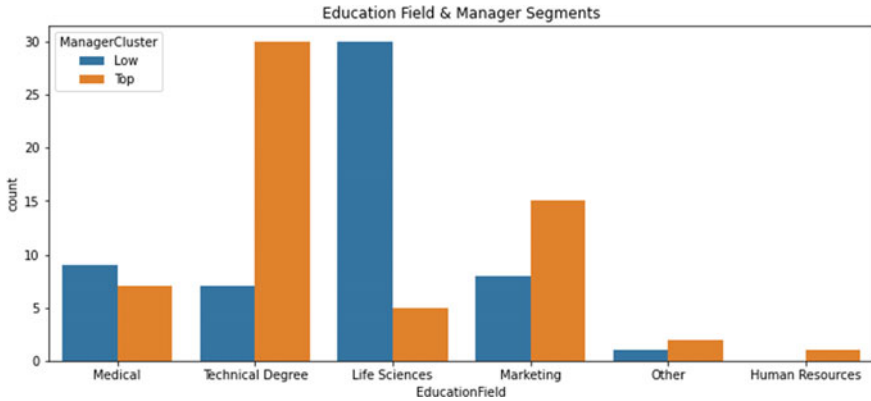


Fig. 4 The performance difference between education degrees

are below 40, have a tenure less than 5, have a total experience of more than 10, have a technical degree in IT and marketing in operations.

5 Training Analytics

Training and development have a significant role in the professional and individual improvement of employees in terms of knowledge, skill, and attitude [42, 43]. As a long-term process aiming to enhance employee performance considering the possible requirements of the job which may arise in the future, the development covers strategic level improvements [43].

Effective training and development support the motivation [44] and self-fulfillment of employees and in turn, reduce employee turnover, ensure successful and sustainable employee retention strategies [45], and contribute to the performance, growth, and competitive power of organizations [46].

Similar to other HR functions, data-based methods and algorithms are used in examining the training and development activities. To automatize the assessment of employees' learning types, the determination of training materials and the development of training strategies, an artificial intelligence-based expert system consisting of a rule-based approach and association rule mining is proposed [47]. This expert system operates using employee information such as department, seniority, and position.

Classifying employees based on their knowledge and expertise also supports HR to determine the training needs. K-means clustering is used to classify the expertise of employees (four classes from excellent to poor) and four classification algorithms including AdaBoost, random forest, support vector machines, and logistic regression are used to identify the development potentials where the random forest performs better with 0.80 accuracy [48]. Employees are classified whether being a part of

Table 7 Training development management-related studies

Purpose	Methods	accuracy	F-measure	Dataset	<i>N</i>	References
A training plan based on performance assessment	k-means clustering	–	–	Real data	100	[49]
Classification	Random forest	0.963	–	Kaggle dataset	54,808	[43]
Development and innovation performance	Logistic regression	–	–	Real data	1384	[46]
Classification	Random forest	0.800	0.796	Real data	1,652,815	[48]

personalized training or not using decision tree, random forest, and support vector machines where the random forest performs better with 0.963 accuracy [43]. This study uses a dataset consisting of variables including department, length of service, the number of previous training, average training score, and awards.

Besides, the department and the industry organizations operate in [43, 46], the tools of training are considered in the studies examining the relationship between training and development and employee performance. Training on the job, coaching, mentoring, using development centers or project teamwork may affect the performance of employees or organizations [46]. Table 7 summarizes development management studies.

Case Study: Data-driven training and development plan through analytics

In this case, an analytics approach was conducted to identify the best training method for the employee to increase performance. The dataset is a modified version of IBM HR Analytics Employee Attrition Data [50]. The main questions of the study are: “Which training method is more effective than others in terms of employee performance? Are there any differences between different departments in terms of training method and performance? Can we predict employee performance through the number of classes and online training?”

The dataset includes performance rating, department information, and training types of 1102 employees of a company. In this case, the main question is to understand which training type is more effective in terms of employee performance. There are multiple factors to evaluate the effectiveness of training. Performance Rating is the observed employee performance. It is the most tangible output of training investments. For this reason, the effect of training on employee performance is explored. Since the training needs of different departments could vary, the relationship between employee performance and training type will be explored on the department level. The list of variables and their definitions can be seen in Table 8.

In the dataset, there are 43 employees from finance, 709 employees from IT, and 350 employees from sales. When correlation analysis is performed, it is seen that

Table 8 List of variables_training and development plan

Variable name	Variable description
OrgFunction	Organizational department of the employee (0: Finance, 1: IT, 2: Sales)
ClassTrainingLastYear	Number of class training assigned to the employee last year
OnlineTrainingsWeeks	Number of online training weeks that employee finished last year
PerformanceRating	Last performance score of the employee

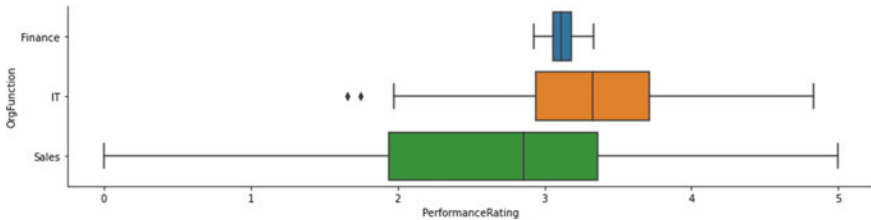


Fig. 5 Boxplot analysis of performance rating

relatively high correlations between performance rating and training types. “Online training weeks” is positively correlated with performance rating with a Pearson coefficient of 0.44. “In Class Trainings” is positively correlated with performance rating with a Pearson coefficient of 0.33.

Boxplot analysis of Class training variable shows that the distributions of three organizations are identical. Secondly, boxplot analysis of Performance Rating given in Fig. 5 shows that IT has a higher average score than Finance and Sales Finance which has a very narrow distribution with low variance. The standard deviation for Finance is 0.09, whereas for IT, it is 0.52 and for Sales it is 0.86.

Exploratory data analysis showed that organizational function is an indicator for performance rating where IT has higher values. To understand the relationship between performance rating and online/class trainings, firstly, the relationship between the target variable performance with other two numeric variables are controlled, and then drilling down to the organizational department level are performed. When the relationship between performance rating and online trainings are explored, a very strong correlation in IT department are occurred. The correlation table on organization department level shows that, Performance scores and online trainings are strongly positively correlated in IT department (0.84). The correlation table between Performance scores and class trainings shows that these two variables are strongly positively correlated in sales department. Since the effect of online and class trainings on employee performance are wanted to be explored, the high correlations in Sales and IT reveal provide a baseline for further analysis. Even though the performance of employees within a company could be related with many factors, the high correlations in Sales and IT enable us to assert that number of class trainings could be a good indicator in Sales department and the number of online trainings that an employee gets could be a good indicator for performance in IT department.

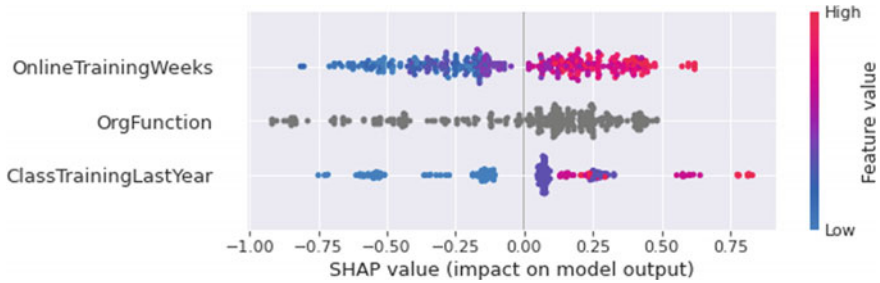


Fig. 6 Shap value_trainin and development management

In the next step, a machine learning algorithm will be applied to predict employee performance through available dependent variables Organizational Function, Online training hours, and the number of class trainings.

To predict employee performance through online and class trainings, a lightgbm regressor model was built. The results show that the model can explain the %78 of target variable (r^2_{score} : 0.78) which is a very good score. Root mean score of the model 0.32 which is also a good error value when it is compared with the mean of target variable performance score (3.1) and standard deviation (0.71). Lastly, shap values of the model were calculated to explain the behavior of the lightgbm model (Fig. 6).

The results of the predictive model show that it is possible to predict employee performance through online and class trainings and organizational department information. Shap values show that higher values of Online trainings also increase the performance score. The exploratory analysis showed this relationship is since Online trainings are highly correlated with performance score in IT department.

In this chapter, the importance of HR analytics is highlighted and showed how HR problems can be solved by using HR analytics tools. HR analytics can be defined as to collect data and transform it into useful information for improving human resource management. HR analytics is very useful for organizations since they create useful insights for managing the organization and help them to reach strategic goals. HR analytics is especially very useful for improving HR functions such as workforce management, recruitment, performance evaluation, and development management. In this chapter, four different HR problems are handled and tried to solve these problems by using HR analytics tools. For future studies, HR analytics tools can be applied to various HR problems.

References

1. Alao D, Adeyemo AB (2013) Analyzing employee attrition using decision tree algorithms. *Comput Inf Syst Dev Inf Allied Res J* 4:17–28
2. Shaw JD, Delery JE, Jenkins GD Jr, Gupta N (1998) An organization-level analysis of voluntary and involuntary turnover. *Acad Manag J* 41:511–525
3. Barrick MR, Mount MK, Strauss JP (1994) Antecedents of involuntary turnover due to a reduction in force. *Pers Psychol* 47:515–535
4. Yadav S, Jain A, Singh D (2018) Early prediction of employee attrition using data mining techniques. In: 2018 IEEE 8th international advance computing conference (IACC). IEEE, pp 349–354
5. Punnoose R, Ajit P (2016) Prediction of employee turnover in organizations using machine learning algorithms. (IJARAI) *Int J Adv Res Artif Intell* 5:22–26
6. Setiawan I, Suprihanto S, Nugraha AC, Hutahaean J (2020) HR analytics: Employee attrition analysis using logistic regression. In: IOP conference series: materials science and engineering. IOP Publishing, p 032001
7. Bhartiya N, Jannu S, Shukla P, Chapaneri R (2019) Employee attrition prediction using classification models. In: 2019 IEEE 5th international conference for convergence in technology (I2CT). IEEE, pp 1–6
8. Ozdemir F, Coskun M, Gezer C, Gungor VC (2020) Assessing Employee Attrition Using Classifications Algorithms. In: Proceedings of the 2020 4th international conference on information system and data mining. Hawaii, USA, pp 118–122
9. Pratt M, Boudhane M, Cakula S (2021) Employee attrition estimation using random forest algorithm. *Baltic J Modern Comput* 9:49–66
10. Sadana P, Munnuru D (2021) Machine Learning Model to Predict Work Force Attrition. In: 2021 6th international conference for convergence in technology (I2CT). IEEE, pp 1–6
11. Mhatre A, Mahalingam A, Narayanan M, et al (2020) Predicting employee attrition along with identifying high risk employees using big data and machine learning. In: 2020 2nd International conference on advances in computing, communication control and networking (ICACCCN). IEEE, pp 269–276
12. Sisodia DS, Vishwakarma S, Pujahari A (2017) Evaluation of machine learning models for employee churn prediction. In: 2017 International conference on inventive computing and informatics (ICICI). IEEE, pp 1016–1020
13. Yahia NB, Hlel J, Colomo-Palacios R (2021) From big data to deep data to support people analytics for employee attrition prediction. *IEEE Access* 9:60447–60458
14. Jain R, Nayyar A (2018) Predicting employee attrition using xgboost machine learning approach. In: 2018 International conference on system modeling & advancement in research trends (SMART). IEEE, pp 113–120
15. Barber AE (1998) Recruiting employees: Individual and organizational perspectives. Sage Publications.
16. Breaugh JA (2013) Employee recruitment. *Annu Rev Psychol* 64:389–416
17. Shahbaz U, Beheshti A, Nobari S, et al (2018) *irecruit: Towards automating the recruitment process*. In: Service research and innovation. Springer, pp 139–152
18. Javed F, Jacob F (2015) Data science and big data analytics at career builder. In: *Big-data analytics and cloud computing*. Springer, London, pp 83–96
19. Martinez-Gil J, Freudenthaler B, Natschläger T (2018) Recommendation of job offers using random forests and support vector machines. In: Proceedings of the workshops of the EDBT/ICDT 2018 joint conference (EDBT/ICDT 2018)
20. Nasar N, Ray S, Umer S, Mohan Pandey H (2020) Design and data analytics of electronic human resource management activities through Internet of Things in an organization. *Softw Pract Exp* 1–17. <https://doi.org/10.1002/spe.3006>
21. Ikram A, Su Q, Fiaz M, Khadim S (2017) Big data in enterprise management: transformation of traditional recruitment strategy. In: 2017 IEEE 2nd international conference on big data analysis (ICBDA). IEEE, pp 414–419.

22. Blau G (1990) Exploring the mediating mechanisms affecting the relationship of recruitment source to employee performance. *J Vocat Behav* 37:303–320
23. Jamil R, Neem H (2013) The impact of outsourcing external recruitment process on the employee commitment and loyalty: empirical evidence from the telecommunication sector of Pakistan. *J Bus Manag* 8:69–75
24. Zottoli MA, Wanous JP (2000) Recruitment source research: current status and future directions. *Hum Resour Manag Rev* 10:353–382
25. Al-Khasawneh AL, Malkawi NM, AlGarni AA (2018) Sources of recruitment at foreign commercial banks in Jordan and their impact on the job performance proficiency. *Banks Bank Syst* 13:12–26
26. Peltokorpi V, Jintae Froese F (2016) Recruitment source practices in foreign and local firms: a comparative study in Japan. *Asia Pac J Hum Resour* 54:421–444
27. Breough JA (2014) Predicting voluntary turnover from job applicant biodata and other applicant information. *Int J Sel Assess* 22:321–332
28. Swamy CJ, Beloor V, Nanjundeswaraswamy TS (2021) Recruitment and selection process in the IT firms. *GIS Sci J* 8:343–356
29. Armstrong M (2006) *Performance management*, 3rd ed. Kogan Page Limited, London and Philadelphia
30. Garg S, Sinha S, Kar AK, Mani M (2021) A review of machine learning applications in human resource management. *Int J Prod Perf Manage*
31. Sahlin J, Angelis J (2019) Performance management systems: reviewing the rise of dynamics and digitalization. *Cogent Bus Manage* 6:1642293
32. Jerath A (2018) The role of analytics in predicting employee performance. In: SHRM. <https://www.shrm.org/shrm-india/pages/the-role-of-analytics-in-predicting-employee-performance.aspx>. Accessed 24 July 2021
33. Keya MS, Emon MU, Akter H, et al (2021) Predicting performance analysis of garments women working status in Bangladesh using machine learning approaches. In: 2021 6th International conference on inventive computation technologies (ICICT). IEEE, pp 602–608
34. Lather AS, Malhotra R, Saloni P, et al (2019) Prediction of employee performance using machine learning techniques. In: Proceedings of the international conference on advanced information science and system. Singapore, pp 1–6
35. Kirimi JM, Moturi CA (2016) Application of data mining classification in employee performance prediction. *Int J Comput Appl* 146:28–35
36. Li MGT, Lazo M, Balan AK, de Goma J (2021) Employee performance prediction using different supervised classifiers. Singapore, pp 6870–6876
37. Pahmi S, Saepudin S, Maesarah N, Solehudin UI (2018) Implementation of CART (classification and regression trees) algorithm for determining factors affecting employee performance. In: 2018 International conference on computing, engineering, and design (ICCED). IEEE, pp 57–62
38. Valle MA, Varas S, Ruz GA (2012) Job performance prediction in a call center using a naive Bayes classifier. *Expert Syst Appl* 39:9939–9945
39. Tang A, Lu T, Lynch Z, et al (2020) Enhancing promotion decisions using classification and network-based methods. In: 2020 Systems and information engineering design symposium (SIEDS). IEEE, pp 1–6
40. Ray AS (2019) AI in performance management. In: peopleHum. <https://www.peoplehum.com/blog/scope-of-ai-in-performance-management> Accessed 24 July 2021
41. Aktepe A, Ersoz S (2012) A quantitative performance evaluation model based on a job satisfaction-performance matrix and application in a manufacturing company. *Int J Ind Eng* 19:264–277
42. Ameen A, Baharom MN (2019) An appraisal of the effect of training on employee performance in an organisation: a theoretical discussion. *Asian J Multidisc Stud* 7:27–31
43. Kaewwiset T, Temdee P, Yooyativong T (2021) Employee classification for personalized professional training using machine learning techniques and SMOTE. In: 2021 Joint international conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunication engineering. IEEE, pp 376–379

44. Ozkeser B (2019) Impact of training on employee motivation in human resources management. *Procedia Comput Sci* 158:802–810
45. Fletcher L, Alfes K, Robinson D (2018) The relationship between perceived training and development and employee retention: the mediating role of work attitudes. *Int J Hum Resour Manage* 29:2701–2728
46. Berber N, Lekovic B (2018) The impact of HR development on innovative performances in central and eastern European countries. *Empl Relat* 40:762–786
47. Chen N, Xu Z, Xia M (2013) Interval-valued hesitant preference relations and their applications to group decision making. *Knowl-Based Syst* 37:528–540
48. Liu J, Huang J, Wang T et al (2021) A data-driven analysis of employee development based on working expertise. *IEEE Trans Comput Soc Syst* 8:410–422
49. Escolar-Jimenez CC, Matsuzaki K, Okada K, Gustilo RC (2019) Enhancing organizational performance through employee training and development using k-means cluster analysis. *Int J Adv Trends Comput Sci Eng* 8:1576–1582
50. <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

Ozgur Akarsu is the head of AI & Data Analytics. He graduated from ITU Industrial Engineering and finished his Ph.D. on HR Analytics in 2016. Ozgur has more than 15 years of experience in digital transformation and analytics in various industries. He also taught classes on machine learning, digital transformation and value chain management in several universities. He is specialized in machine learning and led several projects such as demand forecasting, HR analytics, predictive maintenance.

Cigdem Kadaifci is currently working at the Department of Industrial Engineering, Istanbul Technical University. Dr. Kadaifci got her Ph.D. from Istanbul Technical University. Her research areas are multi-criteria decision making, decision making under uncertainty, scenario analysis, and statistics.

Sezi Cevik Onar is a Professor in the Industrial Engineering Department of Istanbul Technical University (ITU) Management Faculty. She earned her BSc in Industrial Engineering and MSc in Engineering Management, both from ITU. She studied her Ph.D. courses at ITU and Eindhoven Technical University. She has completed her PhD thesis at ITU and visited Copenhagen Business School during these studies. Her Ph.D. was on strategic options. Her research interests include intelligent systems in management, innovation management, creative thinking, strategic management, business planning and multiple criteria decision making. She took part as a researcher in many privately and publicly funded national and international projects including the projects funded by European Union such as intelligent system design, organization design, and human resource management system design. She worked as a visiting professor at Technical University of Troyes. Her refereed articles have appeared in a variety of high impact journals including, *Energy*, *Computers & Industrial Engineering*, *Supply Chain Management: An International Journal* and *Expert Systems with Applications*. Her publications have been cited more than 3000 papers and her h-index is 28.