Optimization of employee turnover through predictive analysis

Stachová Katarína[1[0000-0003-4761-2509]], Baroková Alexandra[2] and Stacho Zdenko[1[0000-0001-5588-8340]]

[1]Institut of management, University of Ss. Cyril and Methodius in Trnava, 917 01 Trnava, Slovakia
katarina.stachova@ucm.sk
zdenko.stacho@ucm.sk
[2] Faculty of Management, Comenius University, Bratislava, Slovakia
alexandra.barokova@fm.uniba.sk

Abstract. In today's technologically driven world, data is essential for decision making process. A comprehensive data analysis is needed to review past practices to improve business performance in the future. The use of data science is found in every prominent domain today and human resources management is one of them. The departure of qualified employees is costly for companies, so it is necessary for them to be able to estimate in advance the probability of their departure and take preventive measures (Jaffara et al., 2019). Once companies have this information, they know how to change their strategy to retain key employees or start hiring new employees in time. For company executives, it is important to link such an analysis of employee departures with the creation of a strategy, and thus obtain cost-effective human resources management. We focused on predicting employee turnover through the Python programming language. Using a correlation matrix that shows the relationships between the various reasons for leaving and decision trees, we provide a management tool to determine the probability of employee turnover. One of the strongest reasons for employee turnover was exhaustion at work, where more than 54% of retired workers worked overtime. Another was income discrepancies between departments. The data analysis points to several differences in the treatment of employees.

Keywords: predictive analysis, human resources management, fluctuation, data analysis, employee turnover

1 Introduction and theoretical background

Human resources management operating in today's dynamic and globalized business environment contributes to increase profits and opportunities to make a positive contribution to overall business performance. Gradual change in the workload of the human resources department plays an important role in organizations and represents an important aspect of the way companies operate (Ruth et al., 2015). The role of human resources specialists is to ensure that the company takes care of its human capital

Management Trends in the Context of Industry 4.0

DOI 10.4108/eai.4-12-2020.2303550

(Kohnová et al. 2020; Stacho et al. 2020, Hitka et al. 2015) and supports the creation and management of procedures, a positive work environment through an effective employee-employer relationship. Today, specialists have more responsibility for human resources management than traditional personnel or administrative tasks. They focus on increasing the value of strategic use of employees and ensuring that employee programs have a positive and measurable impact on business (Soundararajan & Singh, 2016).

The Fourth Industrial Revolution is changing the role of personnel management as we know it and is influencing the current paradigm of talent management (Papula et al. 2019, Jankelová et al. 2017, Lizbetinova et al. 2016). The driving force of the Forth Industrial Revolution is demography (Blštáková et al. 2019, Hitka et al. 2018), technology (Kohnová et al. 2020; Lorincová et al. 2020) and globalization (Kucharčíková et al. 2015; Levický et al. 2013). It is a unique source of opportunities and challenges for human resources to rediscover and renew talent management. Companies support professional practice in talent management by identifying, anticipating, and researching trends that affect talent acquisition and performance (Claus, 2019).

The human resources department gathers lot of information, including employee demographics, recruitment data and performance KPIs, and is therefore also called the data-based department (Rasmussen & Ulrich, 2015). Based on these data, human resources professionals can make less biased recruitment decisions, reduce adverse impacts, and support employees who are more likely to be loyal to the company (Edwards et al., 2019). Recent developments in data collection and analysis tools enable data-based decision-making in all dimensions, including human resources. As a result, human resources analytics is a growing area, therefore we believe it is the right time to elaborate on it. HR analysis is a data-based approach to managing people at work. Many human resources issues can be addressed through a data-based approach. These include decisions such as hiring and retaining employees, performance appraisal, cooperation, and more. If an analyst wants to generate a meaningful overview, it must be rooted in a thorough understanding of the data and the context in which they are collected (Angrave & Charlwood, 2016). The advantage of using a data-based approach is its objectivity and accuracy (Edwards et al., 2019). It is the fourth industrial revolution that requires constant innovation and education, which depends not only on people's abilities, but also on organizational culture. Such a focus can lead to improved analytical capabilities of human resources in organizations and establish a culture of facts and data that many organizations seek to pursue (Heuvel & Bondarouk, 2017).

Predictive analytics use data combined with techniques from mathematics, statistics, and computer science to predict unknown events. The goal of predictive analytics is to provide a good assessment of what could happen to these unknown events. Knowledge of predictive validity is the ability to determine which method has a higher tendency to predict an employee's future behavior. Predictive analysis processes raw data, based on which it points out problematic metrics and emphasizes the critical area to be worked on in strategic decisions (Manuja & Ghosh, 2015). In business analytics, a variable that one wants to predict is marked as a goal. The aim of our analysis is the departure of an employee, i.e. a column in which it is marked whether the employee works in the company or has left. Mondore et al. (2011) argue that the topic of HR analysis has recently been rightly under great pressure. It provides human resources managers with the opportunity to demonstrate the direct impact of their processes and initiatives on business results. Levenson (2011) pointed to the increasing use of HR analysis: “At the beginning of the decade, human resources analysis was not part of the business language. Today, at the end of the decade, Google search will return more than 1.5 million results in the same term."

Based on the above facts in the theoretical part, in the next part of the research, we focused on the use of a specific analytical tool. Using the Python programming language, we create a model for predicting employee turnover. Employee turnover is the process by which employees leave the company. It is essential that the human resources department identifies the factors that retain employees and lead them to leave.

2 Research methodology

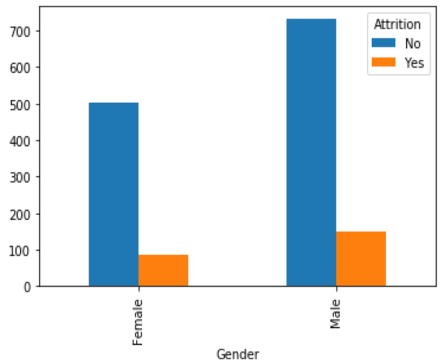
The data sample we work with was created based on IBM's experience and information available in the United States. The database is available and is used for the purpose of creating models later applicable to real data, the accuracy of the data is 85.6% compared to real data. The data set is well organized without missing values, so it can be used to create a correlation matrix showing the relationships between the various reasons for departures and decision trees. We used quantitative methods to identify suitable models for data evaluation and predictive analysis to forecast employee behavior. The research was performed on a hypothetical HR Analytics Employee Attrition & Performance data set, containing 1470 data points (rows) and 35 elements (columns) describing the characteristics of each employee and whether they are still in the company or whether they have left.

At the beginning of the research, a correlation matrix was created, showing the relationship between the individual reasons for employee turnover. After interpreting the results, the strongest correlations were selected and evaluated in more detail using models - decision tree, logistic regression, k-nearest neighbours, and random forest. In order to make an accurate prediction and create an algorithm that can actually be useful, it is common practice in analytics to split data into two samples: training and testing. A training sample of data is used to perform calculations, optimize, and develop the algorithm, and test data are used to validate it (Jaffara et al., 2019). The authors compared the individual models, the most accurate was used to rank the reasons for employee turnover for all employees.

3 Research results

The first step of data analysis is to understand the data sample and its subsequent modification for further tasks. The data set contains 1237 (84% of cases) employees who did not leave the organization, while 233 (16% of cases) left the organization, which considers the data set to be unbalanced because more people remain in the organization than they actually leave. Most models work best when the number of samples in each class is about the same. This is because they are designed to maximize accuracy and reduce errors (Jaffara et al., 2019). Due to the imbalance of this file, it is necessary to approach the interpretation of the results differently and focus on a higher error rate than with the balanced file. In our data sample, the average age of women is 37.33 years and for men 36.6 years, so both distributions are similar. There is a higher number of men in the three departments, but women predominate in the research and development department. Overall, the database was strongly dominated by men (Fig 1).

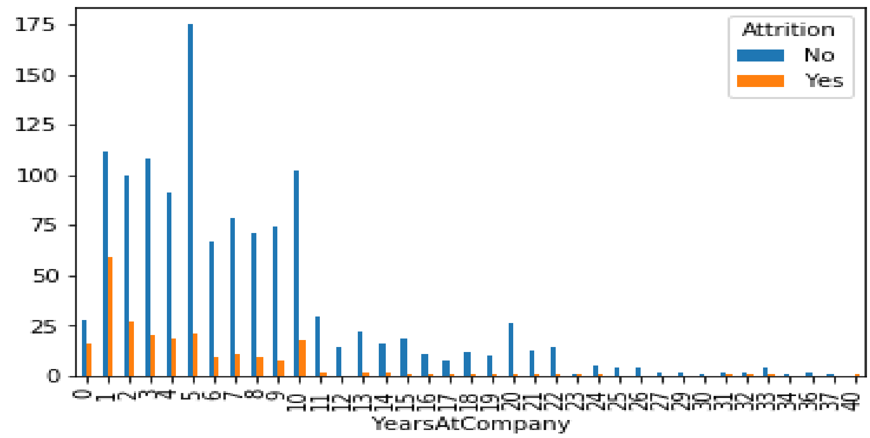
Fig. 1. Distribution of men and women in the database, comparison with employee turnover



Source: Authors analysis of IBM HR Analytics Employee Attrition & Performance dataset

The authors looked at the effect of the length of employment in the company on the number of departures. Most departures were noticed between the first and second year of employment. On the contrary, it was the least in the fifth year of employment (Fig. 2).

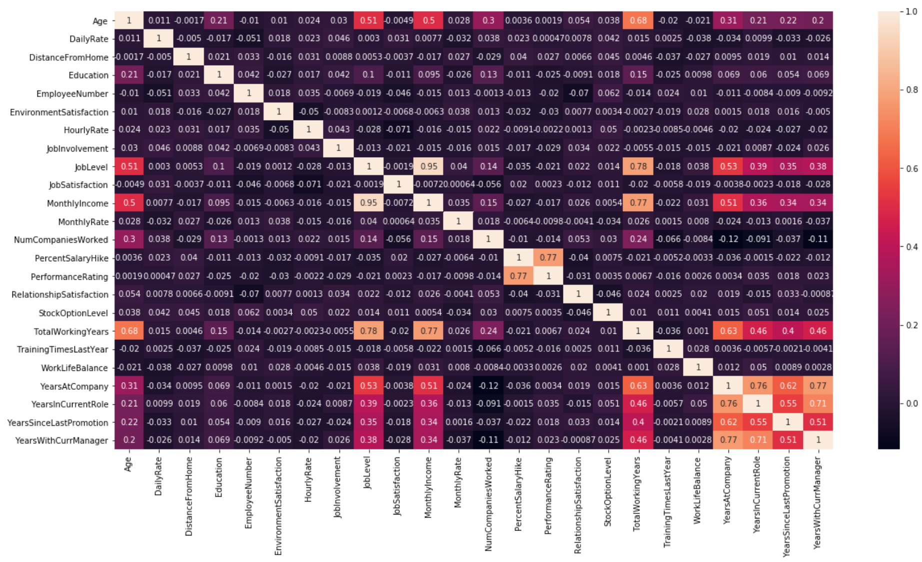
Fig. 2. Length of employment and its comparison with the employee turnover



Source: Authors analysis of IBM HR Analytics Employee Attrition & Performance dataset

The created correlation matrix declares the relationships between the defined reasons for departures. For faster interpretation of the data, the values in the correlation matrix were color-coded, namely: the higher the correlation, the lighter the color (Figure 3).

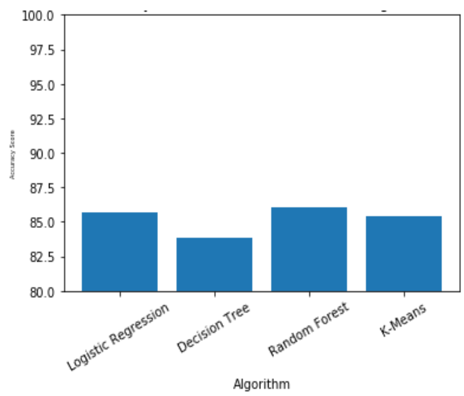
Fig. 3. Correlation matrix of reasons for employee turnover



Source: Authors analysis of IBM HR Analytics Employee Attrition & Performance dataset

We consider the initial analysis of the data through the correlation matrix to be particularly important. Figure 3 shows correlations as: with the growing number of years of the total working year, the average employee income increases, or the better the performance evaluation, the higher the percentage salary increase. The correlation matrix clearly shows the relationships of the reasons for departures, which provides the basis for further analysis. For deeper data analytics, it is necessary to program a model for predicting employee behavior, where the result will be a ranking of reasons for employee turnover according to the number of departures. Data analysis was performed using logistic regression models, decision trees, random forest, and k-nearest neighbours. A comparison of the model accuracy results can be seen in Figure 4. We obtained the model accuracy by creating all the models, where accuracy is automatically calculated.

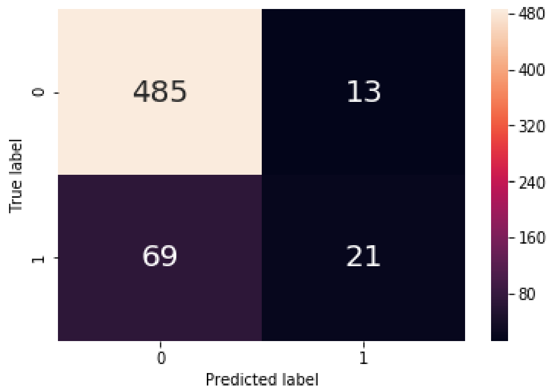
Fig. 4. Accuracy scores for each classification algorithm



Source: Authors analysis of IBM HR Analytics Employee Attrition & Performance dataset

Since the most accurate values were obtained from the random forest model, the authors focused primarily on the interpretation of these results and generated a confusion matrix. The confusion matrix (Fig. 5) is a table with four different combinations of predicted and actual values. Displays horizontal predictions of the number of employees predicted to leave and the number of employees likely to remain in the company. The real numbers of employees are displayed vertically, where 0 means that he will not leave and 1 that the forecast will be fulfilled, and the employee will leave. The matrix works with a training sample of data.

Fig. 5. Confusion matrix for random forest

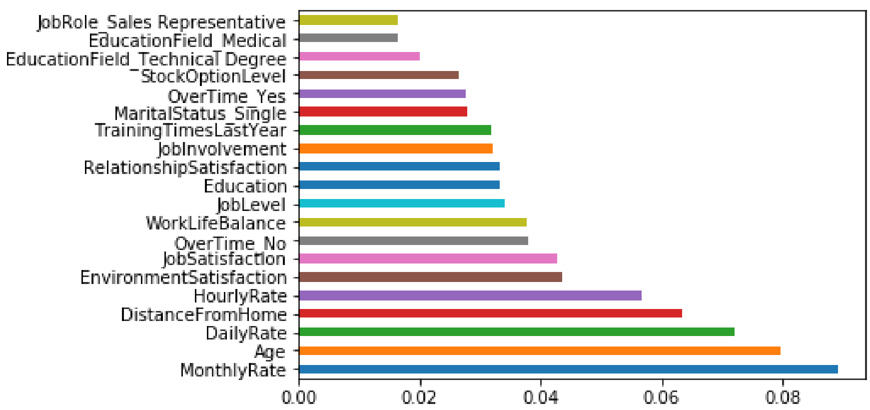


Source: Authors analysis of IBM HR Analytics Employee Attrition & Performance dataset

We predicted that 69 employees would not leave, but in reality, left and 13 employees stayed, even though their departure was predicted. In total, the behavior of 82 employees was incorrectly predicted, which leads to problems, especially for key employees. The degree of accuracy of a classifier is the number of true positives divided by all positive predictions. Low accuracy indicates a high number of false alarms. By creating this matrix, the error rate of the random forest model was found; for the purposes of further research, it is recommended to continue working with the model in the Python program and to program a higher degree of accuracy of the classifier.

The last step in the data analysis is the use of a prepared trained random forest model for the entire database. The result is a graph of ranked reasons for employee departures by employee turnover (Fig 6).

Fig. 6. Sorting the reasons for employee turnover



Source: Authors analysis of IBM HR Analytics Employee Attrition & Performance dataset

The most important factor for employee turnover in this hypothetical organization is monetary because the "Monthly rate" appeared at the top. The dissatisfaction factor was identified in the company's compensation strategy due to the difference in rate between individual departments. The reason may be a bad compensation process or a bad work-life balance. Another important factor seems to be job satisfaction, where the manager's approach and the assigned job are the main factors. Lastly, employee engagement is a critical factor in satisfaction and the organization should constantly involve and motivate employees. Being able to predict whether a certain employee is leaving the company helps to develop a strategy to prevent this situation and gives the company the opportunity to pass this approach on to other employees.

4 Conclusion and discussion

As we get deeper into the Fourth Industrial Revolution, we clearly see changes in human resources management. Digital leaders in human resources monitor emerging human resources technologies and identify and collaborate with the most appropriate technology vendors and platforms for the organization. They are excited to use HR technology to optimize processes and create positive experiences for talent while attenuating risks and possible negative consequences (Bonekamp & Sure, 2015).

The speed of technological change brought by industry 4.0 has created a significant change in human resources management (Stachová et al. 2018, Lorincová et al. 2015, Kirchmayer et al. 2018), which has led to the need to consider new and more effective approaches and tools to predict employee turnover. The goal of using data analysis is to help human resources managers improve the retention rate of valuable employees in the organization, thereby minimizing the cost of employee turnover (Edwards et al., 2019). The success of an organization depends on the right people doing the right things at the right time in the right way (Armstrong, 2007). The authors provide a management tool to determine the likelihood of an employee leaving. HR analyses are gaining attention in organizations that are adopting the digital transformation. The scope has been extended from analyzing the performance of employees' work to providing reports in order to improve organizational processes. Predictive HR analysis has potential that assumes that the accuracy of current and historical data are available. Human resources functions do not face the problem of a lack of available data; on the contrary, they have the problem of having too much data to process properly. The generated matrix points to facts that can be generalized and applied to several companies.

Arulrajah and colleagues (2016) discussed that using qualifiers and classification methods is possible to predict future employee behavior using historical data. The article pointed out the possibility of using data analysis to identify the reasons for employee turnover. The database we worked with was based on information from IBM in the United States. The procedure of predictive data analytics elaborated in the article is applicable for different countries regardless of geographical differences, which occur when interpreting the results. Angrave & Charlwood (2016) describe the importance of understanding the data sample, which was also addressed in the first part of the research. The results of the analysis are objective and will help the human resources department make less biased decisions (Edwards et al., 2019).

Further research could build on and continue by incorporating into the model the possibility of identifying how likely it is for a particular employee to leave and what will be the reason for leaving. As a recommendation, the analytical application can be integrated into the human resources management budgeting application and thereby predict overall profits or savings in the human resources management process, which include recruitment and selection, retention, training and development of new staff, and suggest further steps to be taken. The company can therefore monitor the amount of the budget that it has spent on human resources management and the budget that is to be spent in the future. It takes time and considerable manipulation of data files to ensure that running models match the type of data available.

Acknowledgments: This research was supported by project APVV-17-0656 Transformation of Paradigm in Management of Organizations in the Context of Industry 4.0.

Bibliography

1. Angrave D., Charlwood A. (2016). HR and analytics: why HR is set to fail the big data challenge, Human Resource Management Journal, vol.26(1), pp.1-11.
2. Arulrajah A. A. (2016). Literature review on good governance in the organizations through human resource management: A corporate level analysis, International Business Research, vol. 9 (8), pp. 14
3. Armstrong, M. (2007). Řízení lidských zdrojů, Praha: Grada Publishing
4. Bajzíková, Ľ., Fratričová J., & Kirchmayer, Z. (2019). Základy personálneho manažmentu, Comenius University, Bratislava
5. Bauer T., Erdogan B., Caughlin D.E. (2019). Human Resource Management: People, Data, and Analytics, SAGE Publications
6. Blštáková, J.; Joniaková, Z., Skorková, Z., Némethová, I., Bednár, R. (2019) Causes and Implications of the Applications of the Individualisation Principle in Human Resources Management. AD ALTA: journal of interdisciplinary research – Magnanimitas 9(2), pp. 323-327
7. Bonekamp, L., Sure, M., (2015). Consequences of Industry 4.0 on Human Labour and Work Organisation, Journal of Business and Media Psychology, vol 6(1), pp. 33-40
8. Claus L.(2019). HR disruption-Time already to reinvent talent management. Business Research Quarterly, vol(22), pp.207-215
9. Duan, L., Xiong, Y. (2015). Big data analytics and business analytics, Journal of Management Analytics, vol.2(1), pp. 1-21.
10. Edwards, Kirsten, and Martin R. (2019). Predictive HR Analytics: Mastering the HR Metric, Kogan Page.
11. Fitzenz J., Mattox J. (2014). Predictive Analytics for Human Resources, John Wiley & Sons Inc. Hoboken New Jersey
12. Gypsy N., Rumpam K.S. (2020). Data Science Fundamentals and Practical Approaches: Understand Why Data Science is the Next, BPB Publications, 632s.
13. Hitka, M., Kozubikova, L., Potkany, M. Education and Gender-Based Differences in Employee Motivation. Journal Of Business Economics And Management 9 (1) 80-95 (2018)

14. Hitka, M., Závadská, Z., Jelačić, D., Balážová, Ž. Qualitative indicators of company employee satisfaction and their development in a particular period of time. Drvna in-dustrija: Znanstveni časopis za pitanja drvne tehnologije 66(3), (2015). 235-239. doi.org/10.5552/drind.2015.1420
15. Isson J.P., Harriott J. (2016). People Analytics in the Era of Big Data: Changing the Way You Attract, Acquire, Develop, and Retain Talent. Hoboken, New Jersey, John Wiley & Sons, Inc
16. Jaffara Z., Noorb W., Kanwalc Z. (2019). Predictive Human Resource Analytics Using Data Mining Classification Techniques, International Journal of Computer, vol.32(1), pp. 9-20.
17. Jankelová, N.; Joniaková, Z.; Blštáková, J.; Némethová, I. Readiness of human re-source departments of agricultural enterprises for implementation of the new roles of human resource professionals. - Agricultural economics, 63(10), 461-470, doi:10.17221/189/2016-AGRICECON (2017).
18. Kinney, W. (2017). Python for Data Analysis, O'Reilly UK Ltd
19. Kirchmayer, Z., Fratricova, J. What Motivates Generation Z at Work? Insights into Motivation Drivers of Business Students in Slovakia In: 31st International-Business-Information-Management-Association Milan, ITALY, APR 25-26, (2018)
20. Kohnová, L., Papula, J., Papulová, Z. (2020) Cooperation Models for Employee Education: Analysis on Slovak and Czech Companies. Mobility Internet of Things 2018. Mobility IoT 2018. EAI/Springer Innovations in Communication and Computing. Springer, Cham, ISBN 978-3-030-30910-7, https://doi.org/10.1007/978-3-030-30911-4_22
21. Koubek, J. (2007). Řízení lidských zdrojů, Management Press, 399s.
22. Kucharčíková, A., Tokarčíková, E., Blašková, M.: (2015) Human Capital Management – Aspect of the Human Capital Efficiency in University Education. Procedia - Social and Behavioral Sciences Volume 177, 22 April 2015, Pages 48–60. ISSN 1877-0428 http://www.sciencedirect.com/science/article/pii/S1877042815016869#
23. Levicky, M., Gurcik, L., Porhajas, V., Horvathova, J. 2013. Some aspects of time-demand in the beekeeping production process in Slovakia. Proceedings from VIII. International Conference on Applied Business Research (ICABR 2013), APR 22-26, 2013 East London: SOUTH AFRICA
24. Levit, A. (2019). Humanity Works: Merging Technologies and People for the Workforce of the Future. London, United Kingdom; New York, Ny, Kogan Page Limited
25. Lizbetinova, L., Lorincova, S., Caha, Z. 2016. The Application of the Organizational Culture Assessment Instrument (OCAI) to Logistics Enterprises. Nase More 63(3), 170-176. DOI: 10.17818/NM/2016/SI17
26. Lorincová, S. 2015. The Improvement of the Effectiveness in the Recruitment Process in the Slovak Public Administration. International Scientific Conference: Business Economics and Management (BEM2015), Procedia Economics and Finance, 34, pp. 382-389. DOI: 10.1016/S2212-5671(15)01644-5
27. Lorincová, S., Stachová, K., Stacho, Z., Joniaková, Z., Blštáková, J., Lipoldová, M., Hitka, M. 2020 Defining the differences in corporate culture in wood-processing and forest enterprises BioResources 15(2), pp. 3320-3343
28. Manuja S, Ghosh O. (2015). Strategic Workforce Planning- Impact of HR Metrics and Business Analytics, International Journal of Applied Research, vol.1(4), pp.258-260
29. Mondore S., Douthitt S. a Carson M (2011). Using Human Capital Analytics to Make Excellent Business Decisions, People & Strategy, vol.34 (2), pp.20-27.

30. Papula, J., Kohnová, L., Papulová, Z., Suchoba, M. (2019) Industry 4.0: Preparation of Slovak Companies, the Comparative Study.Smart Technology Trends in Industrial and Business Management. EAI/Springer Innovations in Communication and Computing.Springer, ISBN 978-3-319-76998-1 https://doi.org/10.1007/978-3-319-76998-1_8
31. Rasmussen, T, Ulrich, D. (2015). Learning from practice: how HR analytics avoids becoming a fad. Organizational Dynamics, vol. 44(3), pp.236–242Levenson A (2011). Using Targeted Analysis to Improve Talent Decisions, People & Strategy, vol.34, pp.34-43
32. Ruth, D., Brush, T.H., Ryu, W. (2015). The use of information technology in the provision of HR compensation services and its effect on outsourcing and centralization. Journal of Purchasing and Supply Management pp.25–37
33. Stacho, Z.; Stachova, K.; Caganova, D. Participation of all Employee Categories in Innovation Processes in Slovak Organisations. Mobile Networks and Applications 2020, 1-7, doi: 10.1007/s11036-020-01518-2
34. Stachova, K.; Stacho, Z.; Blstakova, J.; Hlatká, M.; Kapustina, L.M. Motivation of Employees for Creativity as a Form of Support to Manage Innovation Processes in Transportation-Logistics Companies. Naše More 2018, 65(4), 180-186, doi:10.17818/NM/2018/4SI.3
35. Soundararajan R., Singh K. (2016). Winning on HR Analytics: Leveraging Data for Competitive Advantage, SAGE Publications
36. Van del Heuvel S., Bondarouk T.(2017). The rise (and fall?) of HR analytics: A study into the future application, value, structure, and system support, Journal of Organizational Effectiveness: People and Performance, vol.4(2), pp.157-178
37. Zheng, Z. (2015). Introduction to Big Data Analytics and the Special Issue on Big Data Methods and Applications, Journal of Management Analytics, vol.2, pp.281-284