Analysis of Data on Staff Turnover Using Association Rules and Predictive Techniques

Lenka Girmanová, Zuzana Gašparová

ABSTRACT

Purpose: The purpose of this paper is to present the results of an analysis and evaluation of data on employee turnover based on deep data mining using association rules and decision trees in a specific organisation.

Methodology/Approach: For the analysis, we chose deep data mining methods, primarily a search for association rules using the Apriori algorithm in the R programming language. For the sake of supplementation and comparison of results, data were also analysed using the predictive decision trees method, applying the C5.0, rpart and ctree algorithms in the R program.

Findings: The results of the analyses showed that observing the basic principles of correct communication from the beginning of an employment relationship, or during hiring, is justified. Communication and regular conversations between a superior and employees can help identify problems earlier, address them and reduce the number of people leaving the company. The results of the analysis helped the organisation to set measures to reduce the number of an employee leaving.

Research Limitation/implication: A limiting factor in performing such analyses is the availability of quality data in the required quantity. Our most significant advantage when performing our analysis was that quality data were available. To create the final structure of the required data set, we used data from the organisation's internal information systems.

Originality/Value of paper: This contribution offers a new approach to analysing data on employee turnover, whose essence is that we need to find the most interesting and frequent correlations in a significant amount of data.

Category: Case study

1   INTRODUCTION

Organizations nowadays try to ensure the loyalty of their workforce in various ways. However, instead of investing in precise research to understand employees' motivation to leave, or their motivation to stay, most companies invest in additional benefits or measures to search for talent (Zulla Consulting & Partners, 2017).

Employee turnover, or staff turnover, is a measurement of how many employees leave a company (Wilkonson, 2014).

According to Ongori (2007), the essential characteristic of the concept of staff turnover is a situation where an organisation faces the fact that many employees leave because of dissatisfaction, new opportunities on the market, retirement or for other reasons. It is important to distinguish whether employees leave because they are forced to do so or whether they leave voluntarily and to avoid seeing this as an exclusively negative phenomenon. Staff turnover may also have positive aspects. It is mainly in the IT industry that staff turnover is highly beneficial. On the one hand, companies lose their employees, but on the other hand, staff turnover opens the door to other employees, who can bring in new ideas and new ways of thinking (Janice, 2014).

Staff turnover is generally very high when there is an urgent need to hire people in a particular sector or when a large number of people retire. In normal circumstances, employees are likely to leave a company when there are enough opportunities to find jobs elsewhere. The IT industry also struggles with staff turnover due to globalisation, open borders and a high demand for these services and products (Methot, et al., 2017).

Staff turnover as a way of reducing employee numbers is suitable especially in situations where a company needs to reduce costs. According to the expert and speaker John Sullivan (2017), it is true of the vast majority of cases that costs of retaining an employee who is an average performer are lower than costs associated with having the position temporarily unfilled and with replacing him. The most common numerical indicator of staff turnover is the level of staff turnover. This indicator is commonly used in practice even though it is not a very suitable measure, as it does not tell us whether the figure is good or not. As a reference, a healthy level of staff turnover is considered to be 10% (Smith and Rutigliano, 2002). It is not the exact figure that is important, but rather who leaves. Staff turnover is critical for organisations when it involves vital employees and talent. Organizations should, therefore, focus on the level of staff turnover among key employees and those who are top performers. Studies have shown that top performers contribute to the operation of the company ten times more than average performers (Bardessono, 2016). Top companies such as

Microsoft try to maintain their level of employee turnover among top performers, who make up 25% of the total workforce, at 5%, and the average time employees remain at the company is 1.81 years (Peterson, 2017). The level of staff turnover among employees who are the weakest performers should be at least 10% because these employees increase the company's costs.

Every company should specify its level of staff turnover that should be based on strategic planning of the flow of crucial employees throughout the company. It is also necessary to manage the timing of critical employees' leaving. The exact time an employee leaves his team or company helps to understand how significant his leaving is. For example, if an employee who is working on a project decides to leave at the start of the project, this has a lesser impact than a decision to leave the company just before implementing an important part of a project he has been working on (Janice, 2014). The timing of key employees' leaving is an important problem regarding staff turnover even though there are very little literature and very few studies that are dedicated to it. Further research should, therefore, focus on more specific questions such as how to deal with or even how to predict the leaving of key employees' and talent (Janice, 2014). Communication with employees plays an important role here. Even before an employee decides to leave, it is necessary to find out the real reasons for his leaving, which can help prevent key employees' leaving in the future (Sullivan, 2017).

2   METHODOLOGY

Inducing Data Driven culture into an organisation is the key to making fact-based decisions (Mueller, 2017; Zgodavova, Hudec and Palfy, 2017). Markulik, et al. (2018) state that the process has inputs divided into several categories, the most commonly used in practice: people, machines, method, measurement, material, money, market, environment, information, and data collection and archiving are a crucial source of generating information for future decision-making.

For the analysis, we chose deep data mining methods, i.e. the search for association rules and the predictive decision trees method.

The data for analysis come from the company's internal human resources (HR) systems. The company's HR department uses two central information systems. These applications offer a wide range of options to download lists and overviews, which we used to create the final structure of our dataset.

Tab. 1 lists the items used to display the essential statistical indicators and to search for association rules. Information about employees' reasons for leaving is obtained directly from employees using the so-called exit interview.

Table 1 – List of Attributes Used to Analyse Association Rules

Item	Description
Team	the team, the employee, is leaving
Month of Leaving	the month of leaving the company
Duration of service	the number of years at the company
Type of leaving	type of leaving – forced or unforced
Reason for leaving	reason for leaving
Position	employee's position
PPR results	performance and potential review result for the past year
Gender	gender
Age	employee age
Marital status	marital status
Training in the last year	training in the last year – yes or no
Overtime [hours] per month	overtime hours (average hours per month)
On calls [hours] per month	on-call duties (average hours per month)
Home office days per month	home office (average days per month)

Using a different database for the decision trees method was necessary. The list of items is shown in Tab. 2.

Table 2 – List of Attributes Used to Analyse Decision Trees

Item	Description
On-calls [hours] per month	on-call duties (average hours per month)
Home office days per month	home office (average days per month)
Overtime [hours] per month	overtime hours (average hours per month)
Duration of service	the number of years at the company
Age	employee age
Position	employee's position
Exit	employee's exit from the company
Gender	gender
Training in the last year	training in the last year – yes or no
PPR results	the result of performance and potential evaluation for the past year
Change of position last year	change of position last year
Change of CoCe last year	change of location last year
Marital status	marital status

Before we could go ahead with the analysis, we needed to prepare the data thoroughly and clear them of duplicate records. As we worked with two tools, we also created two separate databases containing the data mentioned above. The analysis used data from the past two years, so we needed to unify the values of specific attributes such as the performance and potential review results (PPR) from two different years, where the value was the same, but it was recorded under different symbols. The algorithm would treat it as different values, which would lead to different results and associations in particular. Then we needed to set the right type of individual attributes. In the case of numerical values, we needed to set the type "number" in a ".csv" file, and in the case of a categorical variable, we needed to set the data type "text" for these cells. The data type is significant for the technique of decision trees and association rules.

If a wrong data type was set in the individual analyses, this could show in the results, which would not be correct. Therefore, in the analysis of association rules, it was necessary to change all attributes to the "categorical" data type. The change of data type was performed after entering a command in the R Studio program.

However, when preparing data for analysis through decision trees, numerical values needed to stay numerical and symbolic (text) attributes needed to be changed to categorical.

2.1 Association Rules Mining

According to numerous experts, association rules are seen as one of the most popular methods of data mining. The basic idea is to find the most exciting and most frequent correlations, called associations, in a significant amount of data (Paralič, 2003).

Association rules involve three parameters that need to be explained. These parameters are support, reliability and so-called lift (Agrawal, Imielinski and Swami, 1993).

The result of the algorithm is an association rule, such as one with this form: Training last year = no => Reason for leaving = Career development prospects [0.2, 0.7]. An illustrative example is included in (Tab. 3). Each line represents one transaction and columns represent items.

Table 3 – Illustrative Table for an Association Rule

Duration of service	Type of leaving	PPR results	Training in the last year	Overtime [hours] per month	On-calls [hours] per month	Home office days per month
Four years	Unforced	FIT	Yes	none	none	less than one
Four years	Unforced	FIT	Yes	from 0 to 10	none	less than one

Duration of service	Type of leaving	PPR results	Training in the last year	Overtime [hours] per month	On-calls [hours] per month	Home office days per month
One year	Unforced	IMPROVE	Yes	none	none	less than one
Three years	Unforced	FIT	Yes	from 0 to 10	from 31 to 70	three
Three years	Unforced	FIT	No	none	none	less than one

Support refers to the probability of two items occurring in the same record. The concept of reliability reflects the conditional probability of a transaction containing all items from set Y and at the same time all items from set X. A measure of quality or so-called lift is also used in the case of association rules. If the lift is higher than one, the rule is reliable and of high quality.

The most commonly used and best-known algorithm whose task is to search for association rules from a given set is the Apriori algorithm. This algorithm was also used in our analysis of association rules through the rules package in the R programming language. Apriori searches for association rules in two phases: generation of standard sets of items and generation of the association rules themselves.

Fig. 1 shows a specific association rule from an illustrative set of data that can be interpreted as follows: Employees whose PPR result were Fit and those who had completed training in the previous year and had had no on-calls in a month were all employees who had been at the company for four years. In 40% of the cases, these items appeared together in one record. Lift is 2.5, which indicates a quality rule.

```
      lhs                           rhs                          support confidence lift
[1]  {PPR results=FIT,
      Trainings during last year=yes,
      oncalls [hours] per month=none}  => {Lenght in service=4 years}      0.4        1 2.50
```

Figure 1 – Illustrative Association Rule

2.2 Predictive Analysis of Decision Trees

Our analysis also focused on a different type of data mining, i.e. predictive methods. This type was chosen to supplement the first analysis through association rules, primarily for the sake of comparing results.

The main reason why this data mining technique is useful is its clarity and easy interpretation (Berikov, Litvinenko and Lbov, 2008). The primary aim of this tool is to identify objects represented by the columns in the table regarding classes.

A decision tree is mostly a classifier with a tree structure (Berson, Smith and Thearling, 1995). To create a decision tree, we need to divide the data set into a

testing set and a training set. A testing set specifies a test performed over an instance attribute, where a single branch represents every possible test result. A leaf represents the value of a target property. A decision tree leads from the tree root through individual nodes to the leaf.

Algorithms used to generate decision trees apply the top-down principle. For this analysis, we used algorithms in the R program through the rpart, C5.0 and ctree packages.

2.3   Tool Selection

The R programming language was used for data mining. The environment of this language, the RStudio, met the criteria of availability on the market free of charge and simplicity of user interface. One of the most significant advantages of this language is its rich graphic displaying of results, e.g. box, column, 3D graphs or even more complex interactive graphs. The second significant advantage is the wide range of fields in which it is possible to use the language. Numerous packages have been designed, which are dedicated to topics such as data mining, econometrics or creation of web applications (Berson, Smith and Thearling, 1995).

Creating a code and a procedure for the applied methods of deep data mining was the most critical and challenging part. In the case of procedures, we used knowledge gained from specialist literature and specialist video tutorials freely available on the internet. In the beginning, we needed to load the data into the right format and then process the data to analyse them. Any subsequent procedure is different depending on the type of analysis. Both analyses required the selection of the right R program packages, which allow performing the individual functions, such as the party, party kit, rpart, MASS and c50 packages, etc.

3   RESULTS AND DISCUSSION

It followed from the simple descriptive statistic that the most frequent reasons for leaving were career growth opportunities, dissatisfaction with job description and dissatisfaction with the employee's performance.

	lhs	rhs	support	confidence	lift
[1]	{PPR results=NO PPR, Trainings during last year=no}	=> {Lenght in service=1 year}	0.1488251	0.9268293	3.241786
[2]	{PPR results=NO PPR, Status=single}	=> {Lenght in service=1 year}	0.1618799	0.9253731	3.236693
[3]	{PPR results=NO PPR, Gender=male}	=> {Lenght in service=1 year}	0.1409922	0.9152542	3.201300
[4]	{PPR results=NO PPR, Trainings during last year=no}	=> {Overtimes [hours] per month=none}	0.1449086	0.9024390	1.719573
[5]	{PPR results=NO PPR}	=> {oncalls [hours] per month=none}	0.2140992	0.9590643	1.490149
[6]	{PPR results=NO PPR}	=> {Home office days per month=less than 1}	0.2114883	0.9473684	1.460129
[7]	{overtimes [hours] per month=none}	=> {oncalls [hours] per month=none}	0.4895561	0.9328358	1.449396
[8]	{Trainings during last year=no, oncalls [hours] per month=none}	=> {Home office days per month=less than 1}	0.1788512	0.9319728	1.436401
[9]	{Lenght in service=1 year}	-> {oncalls [hours] per month=none}	0.2637076	0.9223744	1.433142
[10]	{Gender=female, Home office days per month=less than 1}	=> {oncalls [hours] per month=none}	0.1344648	0.9115044	1.416252
[11]	{Trainings during last year=no, overtimes [hours] per month=none}	=> {Home office days per month=less than 1}	0.1592689	0.9172932	1.413776
[12]	{Lenght in service=1 year}	=> {Home office days per month=less than 1}	0.2584856	0.9041096	1.393457
[13]	{Type of leavings=Forced}	=> {Home office days per month=less than 1}	0.1174935	0.9000000	1.387123
[14]	{Lenght in service=2 years}	=> {Trainings during last year=yes}	0.1514360	0.9747899	1.246559
[15]	{Reason of leaving=Career development perspectives, overtimes [hours] per month=from 0 to 10}	=> {Trainings during last year=yes}	0.1109661	0.9444444	1.207754
[16]	{Position=ICT Administrator III}	=> {Gender=male}	0.1227154	0.9400000	1.204080
[17]	{overtimes [hours] per month=from 11 to 25}	=> {Gender=male}	0.1201044	0.9387755	1.202512
[18]	{Lenght in service=4 years}	=> {Trainings during last year=yes}	0.1214099	0.9393939	1.201295
[19]	{overtimes [hours] per month=from 11 to 25}	=> {Trainings during last year=yes}	0.1201044	0.9387755	1.200504
[20]	{Position=ICT Administrator II}	=> {Gender=male}	0.1749347	0.9305556	1.191983

Figure 2 – The Most Robust Association Rules According to the Selected Parameters

3.1 Association Rules

Immediately at the beginning of the analysis of actual data, the set support parameters were 0.1, and the set reliability parameters were 0.9. After removing duplicate association rules, 20 most robust rules were displayed (strong support and reliability). The result of the algorithm according to the selected parameters is shown in Fig. 2.

The 20 most robust rules can be seen in Fig. 2. In the case of support parameters 0.15 and reliability parameters 0.93, the first and the most reliable rule was that an employee without PPR results and training in the past year had been at the company for one year. The first rules are extreme and logical at the same time. Of all employees who leave, it is those who leave within one year that shares the most attributes. Such employees do not participate in PPR reviews, do not complete any training and have no on-call work or work from home. For the sake of a more detailed analysis of the reasons for leaving and to search for more specific association rules, the basic set was divided into so-called subsets according to the reason for leaving. The first was a set with data on employees who had left due to career growth opportunities (Fig. 3).

```
      lhs                                      rhs                                                   support   confidence  lift
[1]   {Overtimes [hours] per month=none,
       Home office days per month=less than 1}  => {Oncalls [hours] per month=none} 0.2391304 0.9649123 1.733827
[2]   {Status=single,
       Overtimes [hours] per month=none}        => {Oncalls [hours] per month=none} 0.2913043 0.9178082 1.649187
[3]   {Overtimes [hours] per month=from 0 to 10,
       Home office days per month=less than 1}  => {Gender=male}                    0.2000000 0.9387755 1.148502
[4]   {Status=single,
       Overtimes [hours] per month=from 0 to 10} => {Gender=male}                   0.2695652 0.9117647 1.115457
[5]   {PPR results=FIT,
       Gender=male}                             => {Trainings during last year=yes} 0.2956522 0.9444444 1.097082
[6]   {Gender=male,
       Overtimes [hours] per month=from 0 to 10} => {Trainings during last year=yes} 0.3347826 0.9390244 1.090786
[7]   {PPR results=FIT,
       Status=single}                           => {Trainings during last year=yes} 0.2434783 0.9333333 1.084175
[8]   {Status=single,
       Overtimes [hours] per month=from 0 to 10} => {Trainings during last year=yes} 0.2739130 0.9264706 1.076203
[9]   {Position=ICT Administrator II,
       Gender=male}                             => {Type of leavings=Unforced}      0.2086957 1.0000000 1.004367
[10]  {Gender=male,
       Status=married}                          => {Type of leavings=Unforced}      0.2043478 1.0000000 1.004367
[11]  {Status=married,
       Trainings during last year=yes}          => {Type of leavings=Unforced}      0.2173913 1.0000000 1.004367
[12]  {PPR results=FIT,
       Status=single}                           => {Type of leavings=Unforced}      0.2608696 1.0000000 1.004367
[13]  {PPR results=FIT,
       Gender=male}                             => {Type of leavings=Unforced}      0.3130435 1.0000000 1.004367
[14]  {PPR results=FIT,
       Trainings during last year=yes}          => {Type of leavings=Unforced}      0.3565217 1.0000000 1.004367
[15]  {Overtimes [hours] per month=from 0 to 10,
       Home office days per month=less than 1}  => {Type of leavings=Unforced}      0.2130435 1.0000000 1.004367
[16]  {Status=single,
       Overtimes [hours] per month=from 0 to 10} => {Type of leavings=Unforced}     0.2956522 1.0000000 1.004367
[17]  {Gender=male,
       Overtimes [hours] per month=from 0 to 10} => {Type of leavings=Unforced}     0.3565217 1.0000000 1.004367
[18]  {Trainings during last year=yes,
       Overtimes [hours] per month=from 0 to 10} => {Type of leavings=Unforced}     0.3695652 1.0000000 1.004367
[19]  {Gender=male,
       Overtimes [hours] per month=none}        => {Type of leavings=Unforced}      0.2782609 1.0000000 1.004367
[20]  {Gender=male,
       Oncalls [hours] per month=none}          => {Type of leavings=Unforced}      0.3913043 1.0000000 1.004367
```

Figure 3 – Association Rules – Career Growth Opportunities I

As we can see from the Fig. 3, the first and the second rules contain the on-calls and overtime items. In both cases, the values of these items were zero, so employees who state career growth opportunities as their reason for leaving had no on-calls or overtime. Based on this result, it is possible to conclude that on-calls and overtime had no impact on career growth.

We also focused on this reason for leaving in the case of employees who had been at the company for one year. The results of running the algorithm after adding the "duration of the service=1 year" are shown in Fig. 4.

```
     lhs                                      rhs                                              support   confidence  lift
[1]  {PPR results=NO PPR}                     => {Lenght in service=1 year} 0.2325581 1.0000000 3.307692
[2]  {Trainings during last year=no,
      Overtimes [hours] per month=none,
      Home office days per month=less than 1}  => {Lenght in service=1 year} 0.1162791 0.9090909 3.006993
```

Figure 4 – Association Rules – Career Growth Opportunities II

Employees who had been at the company for one year and had not participated in PPR reviews had not had any training, home office or overtime. We could conclude based on this association rule that PPR reviews are critical as they give employees an opportunity to define their career paths at the company, plan their training and obtain their superiors' evaluation.

Since education has an impact on careers, the parameters for searching for association rules were extended to include a rule for searching for associations in the case of those employees who had completed training in the previous year

(Fig. 5). The results implied that employees who had been at the company for four years had completed some training in the previous year. This rule shows that it is likely that the company does not offer an employee who has been at the company for quite large opportunities for career growth even through training.

Figure 5 – Association Rules – Career Growth Opportunities III

Rule number 11 is also impressive, as employees who were 30 years of age and completed training stated career growth opportunities as their reason for leaving. The group of employees aged 30 is probably a critical one, as they tend to reconsider their careers and change their work environments if they can.

Another frequent reason for leaving was "dissatisfaction with job description". The interpretation of the strongest rule is that employees who had been at the company for one year had had no overtime (Fig. 6). Other rules suggest that employees who had been at the company for one year were single and had worked from home less than one day a month stated dissatisfaction with the job description as their reason for leaving. As we can also see in other associations, several rules contain the "NO PPR" item, which means that the given employees had not had a review. These facts allow concluding that it is essential to evaluate employees.

Figure 6 – Association Rules – Dissatisfaction With Job Description I

As the second most robust rule contained the marital status = single item, we focused on the employees who are single (Fig. 7). It is single employees aged 25 and 26 that left the company due to dissatisfaction with job description with 100% reliability. 25-26 is the age when most university graduates find their first jobs. If they did not work during their studies or had no experience with work

discipline, they may surprise by the fast pace of work, and because of having unrealistic expectations, they state dissatisfaction with the job description as their reason for leaving. This result prompted the inclusion of another question in the exit interview. The question was whether this was the employee's first work experience.

	lhs	rhs	support	confidence	lift
[1]	{Age=25}	=> {Status=single}	0.1046512	1.0000000	1.228571
[2]	{Oncalls [hours] per month=from 31 to 70}	=> {Status=single}	0.1162791	1.0000000	1.228571
[3]	{Age=26}	=> {Status=single}	0.1395349	1.0000000	1.228571

Figure 7 – Association Rules – Dissatisfaction with Job Description II

The next most frequent reason for leaving was "work-life balance" (Fig.8). The most robust rules show that those employees who stated this reason for leaving were single and had had no overtime or on-calls. The second rule was that these were cases of unforced leaving of employees who had been at the company for one year and their average use of benefits had been less than one day a month.

	lhs	rhs	support	confidence	lift
[1]	{Status=single, Overtimes [hours] per month=none}	=> {Oncalls [hours] per month=none}	0.1363636	1.0	4.400000
[2]	{Lenght in service=1 year, Type of leavings=Unforced}	=> {Home office days per month=less than 1}	0.1136364	1.0	1.760000
[3]	{Lenght in service=2 years, Position=ICT Administrator II}	=> {Home office days per month=less than 1}	0.1136364	1.0	1.760000
[4]	{Status=single, oncalls [hours] per month=from 31 to 70}	=> {Home office days per month=less than 1}	0.1136364	1.0	1.760000
[5]	{Gender=male, Overtimes [hours] per month=from 0 to 10, oncalls [hours] per month=from 31 to 70}	=> {Home office days per month=less than 1}	0.1363636	1.0	1.760000
[6]	{Status=single, Overtimes [hours] per month=from 0 to 10}	=> {Home office days per month=less than 1}	0.2045455	0.9	1.584000
[7]	{Lenght in service=1 year, Home office days per month=less than 1}	=> {Status=single}	0.1363636	1.0	1.466667
[8]	{Lenght in service=1 year, Type of leavings=Unforced}	=> {Status=single}	0.1136364	1.0	1.466667
[9]	{Position=ICT Administrator II, PPR results=FIT, Home office days per month=less than 1}	=> {Status=single}	0.1136364	1.0	1.466667
[10]	{Gender=male, Overtimes [hours] per month=from 11 to 25}	=> {Status=single}	0.2045455	0.9	1.320000

Figure 8 – Association Rules – A Work-Life Balance I

Fig. 9 shows the results of an analysis of association rules where the search was extended to include the parameter of searching for rules for single employees. Again, the rule emerging here is that these employees had been at the company for one year, had had no overtime or on-calls. In most cases these employees' position was ICT Administrator II, their PPR result was Fit, and their average use of benefits had been less than one day a month.

	lhs	rhs	support	confidence	lift
[1]	{Lenght in service=1 year}	=> {Status=single}	0.1363636	1.0000000	1.466667
[2]	{Gender=male, Overtimes [hours] per month=none, oncalls [hours] per month=none}	=> {Status=single}	0.1136364	1.0000000	1.466667
[3]	{Position=ICT Administrator II, PPR results=FIT, Home office days per month=less than 1}	=> {Status=single}	0.1136364	1.0000000	1.466667

Figure 9 – Association Rules – Work-Life Balance II

3.2 Decision Trees

```
Lenght in service <= 1: no (789/57)
Lenght in service > 1:
:...PPR results in {Best fit,Fit,Grow,Improve,Jump,Move}: no (3198/229)
    PPR results = NO PPR:
    :...Training last year = no:
        :...Lenght in service <= 1.8: yes (5)
        :   Lenght in service > 1.8:
        :   :...Status in {divorced,married,widow}: no (36/4)
        :       Status = single:
        :       :...Position in {EG1,EG2}: yes (8/2)
        :           Position in {EG3,EG4,MG,N/A}: no (4)
        Training last year = yes:
        :...Status = widow: yes (0)
            Status = divorced: no (3)
            Status in {married,single}:
            :...Home office <= 0.3333333:
                :...Position in {EG1,EG2,EG3,EG4,N/A}: yes (113/9)
                :   Position = MG: no (12/5)
                Home office > 0.3333333:
                :...Overtime <= 12.375: no (62/21)
                    Overtime > 12.375: yes (7)
```

Figure 10 – Decision Tree of the C5.0 Algorithm

In this phase of the analysis, we used algorithms in the R program through the C5.0, rpart and ctree packages.

Fig. 10 shows an output from the R program after entering the command for generating the decision tree through the C5.0 algorithm. The primary classifier from which the tree was developed was the Exit variable, i.e. whether the given employee had left the company. The end leaves of this decision tree should result in information about whether an employee has left the company or not. The first branch where the tree branches out is the "Duration of service" attribute, i.e. how long an employee has been at the company. This branch branches out further depending on whether an employee has been at the company for more than one year. If not, the algorithm does not continue, as only 57 employees ended their employment, which is too low a number for further development of the tree. The tree then continues to branch out according to PPR results. Here we can see another end leaf, where the tree stops developing if the employees participated in the performance and possible review. Of these employees, 229 terminated their employment. If the employees did not participate in the PPR review, the tree continues and the attribute of whether employees had training in the previous year. According to this attribute, the tree branches out into two separate parts.

If the employees did not participate in training in the previous year, the algorithm continues with the "Duration of service" attribute, like in the first layer. If this was less than 1.8 years, then five employees left the company, to which the specifications mentioned above also applied. If the employees had been at the company longer than 1.8 years, were single and in a junior position, they left the company.
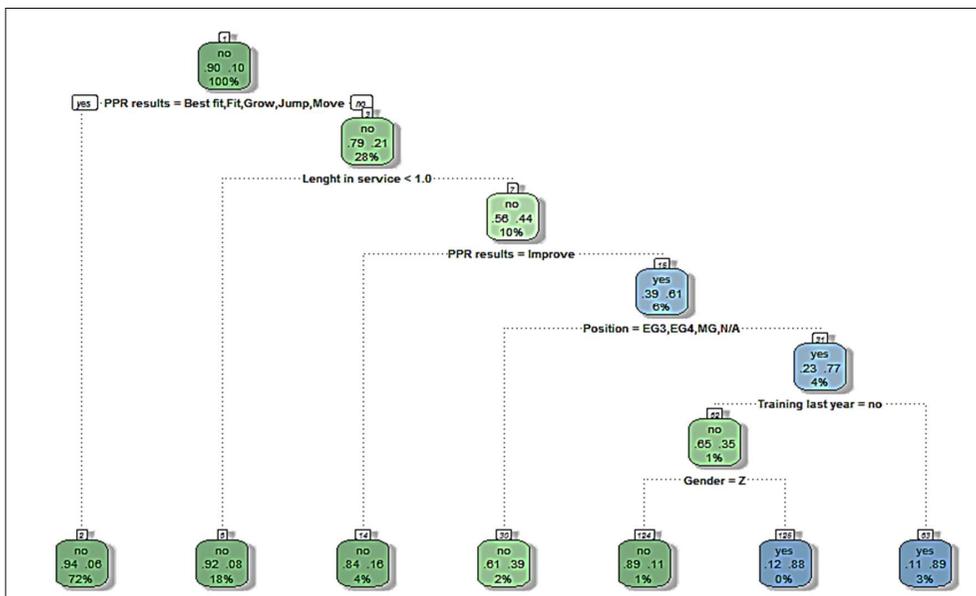
Figure 11 – Decision Tree of the rpart Algorithm

The second part of this tree applies to those employees who had had training in the previous year. If they were single or married and their average use of benefits had been less than 0.33 days a month, 113 employees terminated their employment. If they had worked an average of more than 0.33 days a month from home and had not worked more than 12.375 days of overtime on average, they stayed at the company and did not terminate their employment.

Another algorithm we used was rpart. The result of this algorithm is a decision tree shown in Fig. 11. The different colours make it easier to find the required result. The green colour refers to nodes, or leaves, where the employees did not leave the company, and the blue colour refers to those where they did.

The tree begins with the attribute of PPR results. If the employees' results were Bes fit, Fit, Grow, Jump and Move, they did not leave the company in 72% of cases. The remaining 28 % of employees had a different result, i.e. Improve or no result if they had not participated in the review. If the employees had been at the company more than one year and their PPR result was other than Improve, i.e. they had not participated in PPR, they terminated their employment (the first blue leaf). The second blue leaf shows that employees who had been in EG1 and EG2 positions left the company. The last leaves show that the employees who leave the company do not participate in training and if they do, they turn out to be women who decide to leave the company.

The last algorithm we used in our decision tree analysis was the ctree algorithm. For ease of understanding, Fig. 12 shows a decision tree in the graphic form of a standard decision tree. The end of this decision tree features rectangles that allow us to compare when the employee for the given branch left his employment and

when he did not. The results are also shown as a percentage ratio. The darker the part in the broader area of the rectangle, or the higher the value it reaches, the more important and more interesting this branch is for prediction.
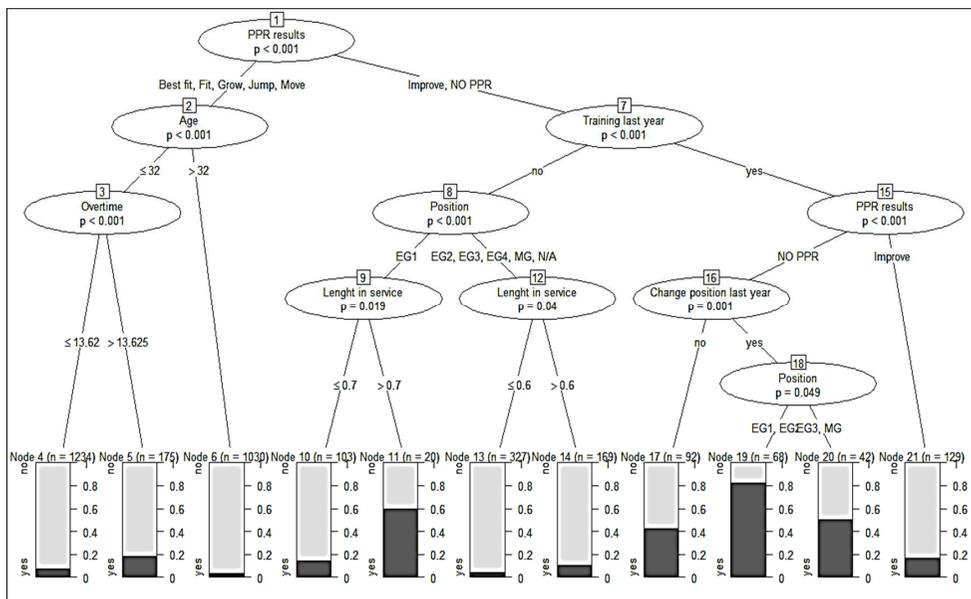

Figure 12 – Decision Tree of the ctree Algorithm I

The root of the tree contains employees with PPR results in Best fit, Fit, Grow, Jump a Move. This part branches out further according to age. If an employee was aged 32 or less and had worked an average of more than 13.625 hours of overtime a month, almost 20% of these employees, who share the given attributes, out of the total 175 employees left the company, see Leaf no. 5.

The right part of the decision tree applies to employees whose PPR result was Improve or had not participated in the review. If these employees had not had training in the previous year, their positions were at the EG1 level and had been at the company more than 0.7 years, then in 60% of cases, these employees left the company, see Leaf no. 11.

Employees who had done training in the previous year, but had not participated in the PPR review and had not changed their position in the previous year, then 40% of the total of 92 employees left the company, see Leaf no. 17 Even if they had changed their position and stayed at the EG1 level, they left the company in 80% of cases. Leaf no. 19 since in most cases the root of the tree contained PPR results, this column was removed from the database, and the algorithm was rerun (Fig. 13).
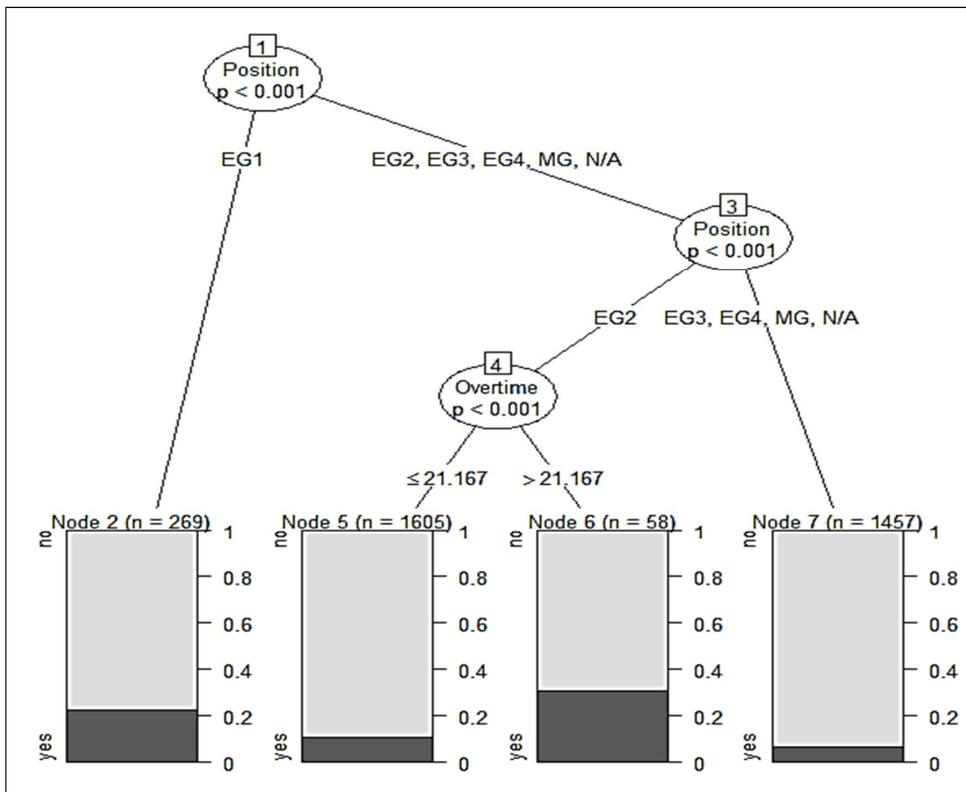
Figure 13 – Decision Tree of the ctree Algorithm II

Compared to the above decision tree, this decision tree is much smaller and only contains four leaves. The root of this tree contains position. If an employee's position is EG1, then 20% of the total of 269 employees left the company. In the case of EG2 positions, a branch for overtime was added to the decision tree. If employees' average overtime per month is more than 21.167, around 30% of them leave the company.

4   CONCLUSION

We first performed a simple analysis of the frequency of the individual types of exit. That served as preparation for further analyses and provided results that were not negligible. The most frequent reasons for leaving the given company in the past two years were career growth opportunities, dissatisfaction with job description and dissatisfaction with the employee's performance. These results were also confirmed in analyses using association rules and decision trees.

The results of the analyses generally showed that observing the basic principles of correct communication from the beginning of an employment relationship, or during hiring, is justified. It is necessary for every potential employee to be familiar with their job description, so they do not leave when they are still in

their trial period, which is a point when employees often state dissatisfaction with the job description as their reason for leaving. Communication and regular conversations between a superior and his employees can help identify problems earlier, address them and reduce the number of people leaving the company. The following recommendations were put forward for the company:

- Improve the methods of graduate hiring, i.e. prepare future employees by offering internships to students who can switch to full-time employment after graduation.

- At the end of the trial period, discuss with the given employees whether the position for which they have been hired meets the requirements and expectations they had before they started.

- Set an education plan and training plans for the individual types of employees.

- Motivate loyal employees with benefits.

- Monitor the amount of overtime and if an absolute limit is exceeded, review the employees' job description to avoid unnecessary overtime that may have an adverse effect on them.

The most significant advantage when performing the analysis was the availability of quality data in large amounts. It was not possible to perform such an analysis in specific industries or at certain companies precisely because detailed databases and overviews were not available.

ACKNOWLEDGEMENTS

REFERENCES

Agrawal, R., Imielinski, T. and Swami, A., 1993. Mining Association Rules Between Sets of Items in Large Databases. In: P. Buneman and S. Jajaodia eds., *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*. Washington, DC, USA, 25-28 May 1993. New York, USA: ACM.

Bardessono, D., 2016. 5 Signs You're A "Unicorn" Employee, Benchmark Consulting, [online] 10 October. Available at: <http://www.benchmarkhr.com/5-signs-youre-a-unicorn-employee/> [Accessed 05 March 2018].

Berikov, V., Litvinenko, A. and Lbov, G.S., 2008. *Methods for statistical data analysis with decision trees*. [pdf] s.l.: Sobolev Institute of Mathematics. Available at: <http://www.math.nsc.ru/AP/datamine/eng/context.pdf> [Accessed 05 March 2017].

Berson, A., Smith, S. and Thearling, K., 1995. *An Overview of Data Mining Techniques*. [pdf] s.n. Available at: <http://weber.itn.liu.se/~jimjo94/courses/TNM048/documents/DM-Techniques.pdf> [Accessed 05 March 2017].

Janice, L., 2014. New Perspectives on Staff Turnover in the IT Field. *Knowledge HEC*, [online] 11 December. Available at: <http://www.hec.edu/Knowledge/Strategy-Management/Human-Resources-Management/New-Perspectives-on-Staff-Turnover-in-the-IT-Field> [Accessed 08 January 2017].

Markulik, Š., Cehlár, M. and Kozel, R., 2018. Process approach in the mining conditions. *Acta Montanistica Slovaca*, 23(1), pp.46-52.

Methot, J.R., Lepak, D., Shipp, A.J. and Boswell, W.R., 2017. Good Citizen Interrupted: Calibrating a Temporal Theory of Citizenship Behavior. *ACAD MANAGE REV*, [e-journal] 42(1), pp. 10-31. http://dx.doi.org/10.5465/amr.2014.0415.

Mueller, R., 2017. How Can An Organization Transform Its Culture To Become More Data Driven?. MacroSoft, [online] 13 October. Available at: <https://www.macrosoftinc.com/blog/how-organization-transform-culture-more-data-driven.html> [Accessed 02 March 2018].

Ongori, H., 2007. A review of the literature on employee turnover. African Journal of Business Management, [e-journal] 1(2), pp.46-54.

Paralič, J., 2003. *Discovering knowledge in databases*. Košice: Elfa.

Peterson, B., 2017. Travis Kalanick lasted in his role for 6.5 years — five times longer than the average Uber employee. *Business Insider,* [online] 20 August. Available at: <http://www.businessinsider.com/employee-retention-rate-top-tech-companies-2017-8> [Accessed 05 March 2018].

Smith B. and Rutigliano T., 2002. The truth about turnover. *Business Journal*, [online]. Available at: <http://www.gallup.com/businessjournal/316/truth-about-turnover.aspx> [Accessed 16 January 2017].

Sullivan, J., 2017. The Ideal Turnover Rate. *Monster*, [online]. Available at: <http://hiring.monster.ca/hr/hr-best-practices/recruiting-hiring-advice/strategic-workforce-planning/employee-turnover-rate-canada.aspx> [Accessed 15 February 2017].

Wilkonson, J., 2014. Employee Turnover Definition. *The Strategic CFO*, [online] 7 April. Available at: <https://strategiccfo.com/employee-turnover/> [Accessed 05 March 2018].

Zgodavova, K., Hudec, O. and Palfy, P., 2017. Culture of quality: insight into foreign organisations in Slovakia. *Total Quality Management & Business Excellence*, [e-journal] 28(9-10), pp.1054-1075. http://dx.doi.org/10.1080/14783363.2017.1309120.

Zulla Consulting & Partners, 2017. Should I stay or should I go – Why your employees have this doubt?. *Zulla Consulting & Partners*. [online] 26 April. Available at: <http://zulla-consulting.com/employee-fluctuation> [Accessed 15 February 2018].

ABOUT AUTHORS

Ing. Lenka Girmanová, PhD. – Lecturer, Technical University of Košice, Faculty of Materials, Metallurgy and Recycling, Institute of Materials and Quality Engineering, Department of Integrated Management Systems, e-mail: lenka.girmanova@tuke.sk.

Ing. Zuzana Gašparová – Graduate, Technical University of Košice, Faculty of Materials, Metallurgy and Recycling, Institute of Materials and Quality Engineering, Department of Integrated Management Systems, e-mail: zuzana.herpak@gmail.com.