26-osios Lietuvos jaunųjų mokslininkų konferencijos „Mokslas – Lietuvos ateitis" teminė konferencija
Proceedings of the 26th Conference for Junior Researchers "Science – Future of Lithuania"

EKONOMIKA IR VADYBA / ECONOMICS AND MANAGEMENT

2023 m. spalio 17 d., Vilnius, Lietuva
17 October 2023, Vilnius, Lithuania

eISSN 2029-7149
eISBN 978-609-476-343-4
Article Number: vvf.2023.006

Šiuolaikinio verslo aktualijos / Actualities of Modern Business

https://vilniustech.lt/jmk-vvf

IMPROVING PROCESS PERFORMANCE MANAGEMENT BY PREDICTING EMPLOYEE ATTRITION IN INTERNATIONAL COMPANY

Aleksei IURASOV*, Virginija BARGAILĖ

Department of Business Technologies and Entrepreneurship, Faculty of Business Management, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania
**E-mail: aleksei.iurasov@vilniustech.lt*

Received 29 December 2022; accepted 27 March 2023

Abstract. Employee attrition has become significant problem, because it affects organisation's process performance, and generates both – financial and business management loss. There is a tendency of growing employee attrition rate, and this is a reason why solution should be taken to lower it. It might be possible to lower employee attrition rate and thus to find a solution to this issue by predicting employee attrition and taking proactive actions for potential leavers. For employee attrition prediction, machine learning techniques and algorithms might be used. Machine learning algorithms makes predictions, using information learned from historical data, for this reason those predictions are more accurate, than just intuition-based predictions. There exist variety of different machine learning algorithms, hence it should be chosen depending on its reliability. Decision Tree algorithm was discovered to be the most reliable (97.4% accuracy) for predictive model in this study. However, having the predictive model itself doesn't mean, that employee attrition problem will be solved. Model, as a tool should be integrated into organisation's business process management, to ensure continuity. The appropriate framework ensuring this continuity could be process performance management, which enables process measurement, control, and it helps to achieve better results in an organisation by setting up KPIs. KPIs is the crucial part of process performance management because it aligns business activities with strategy. Therefore, a strategic employee attrition solution should be connected to one of organisation's KPI, to reach the target and ensure continuity.

Keywords: process performance, key performance indicators, attrition prediction, machine learning.

Introduction

Employees are leaving their works daily, for any varied reasons, hence it looks it should be very natural process. But according to Jain et al. (2020), the employee attrition is relevant and ongoing problem internationally, thus it causes loses of valuable and experienced employees. Employee, leaving the company, takes away "know-how", which is the main business advantage. For this reason, to become superior against the competitors, knowledge and experience-based organisation should put effort in minimizing employee attrition by keeping the experienced employees in the organisation (Alao & Adeyemo, 2013; Haldorai et al., 2019). Moreover, employee attrition might affect productivity, and planning continuity in the company (Yahia et al., 2021), because company might lose a productive employee as well (Al-Darraji et al., 2021). Based on statistics, employee attrition rate in the year 2021 was 57.3%, and the average cost per new hire was $4129 (Raza et al., 2022). Recruiting new employee requires human resource (HR) effort, training, development, and integration into the unfamiliar environment (Al-Darraji et al., 2021). Therefore, this is massive loss which makes impact for any kind of organisation (Raza et al., 2022). However, this problem might be solved or reduced, by predicting employee attrition (Al-Darraji et al., 2021), using data science and big data analytics (Yahia et al., 2021). This solution would assist businesses and their HR managers in taking influencing factors into account and changing how they attract new and retain existing employees (Yahia et al., 2021).

The aim of this research: to suggest decision support model, which would help to predict employee attrition and provide the recommendations on how to reduce employee attrition rate.

This study's tasks are as follows:

- First, perform literature analysis of employee attrition problem relevance to process performance management framework.

– Second, review the existing literature of potential solutions used to solve business problems, and analyse earlier studies specifically related to the employee attrition.
– And third, propose a model with an algorithm that can predict employee attrition. Following that, recommendations on company behaviour should be provided as well, which may help to reduce employee attrition ensuring continuity.

1. Employee attrition problem relevance to process performance management framework

Process Performance Management (PPM) is a structured process for improvement, which entails managing the overall performance of the whole company and helps to achieve the end state incorporating the identification and satisfaction of needs (Hey, 2017). Moreover, Gruman and Saks (2011) emphasis, that PPM is the concept, through which work is accomplished, thus it is crucial for company's efficiency. In fact, PPM enables monitor and control of the process, which is inseparable part to manage business process performance properly (Balaban et al., 2011; Ongena & Ravesteyn, 2020). To be able to monitor the process, appropriate performance measurement and assessment are required. As a result, Key Performance Indicators (KPI) should be identified and established in the company (Hey, 2017; Ongena & Ravesteyn, 2020). Difficulty by setting up KPIs might differ, according to availability of information. Moreover, after setting up the KPIs, processes should always be improved in case if it doesn't meet settled targets anymore (van der Aalst et al., 2016). The below is provided basic recommendations on how to establish KPIs:

Fleischmann (2012) advises setting relevant KPIs that are aligned with the organisation's strategy; as a result, KPIs will benefit the organisation. Additionally, author suggests consultations and experience exchange between the organisations before the setting up new target values, thus it is often difficult to determine them realistic.

Folan and Browne (2005) by establishing new target values suggests involving all organisation's employees – from regular employees to top management. Target values might be established during a series of workshops, brainstorming sessions, interviews, and even pilot process models. Those target values should be added into the feedback loop, which would link them to manager and employee performance appraisals. An organisation must ensure that those values are relevant to managers and employees in performing their day-to-day jobs.

Having set an appropriate performance value, it is possible to make a comparison between the planned and achieved KPI's. Discrepancies between what is planned and achieved will result in implementing measures contributing to planning performance enhancement. Van der Aalst et al. (2016) defines KPIs all over the different performance angles such as time, quality, cost, flexibility, etc. Moreover, unified intelligence approach is necessary, to be able to have real-time status of process performance and to coordinate organisation's assets, related to Business Process Management, e.g.: business processes, processes related data and employee's information (Balaban et al., 2011).

Nowadays, the business environment becomes more competitive (Peters, 2019), and it could become an important issue in both, national and global business (Brenes, 2000). However, previous studies have highlighted the importance of companies realizing every available resource when competing with international companies. As a result, it is critical to retain professional employees to maintain a competitive advantage. Besides, competitive advantage must be incorporated into the company's strategy to ensure long-term viability as well (Brenes, 2000). Moreover, clients become more demanding, thus organisations must obtain the best performance from their employees to meet their productivity, customer experience, quality, and profitability targets (Peters, 2019). Phillips et al. (2016) echoes, that organisations, which are keen on to be high performing, should be more innovative, excel at KIPs, should keep key employees and to attract new employees by strengthen employee value propositions (Phillips et al., 2016). Apart from that, it is uneasy to keep loyal employees in the organisation and based on the report of the Bureau of Labor Statistics, employee attrition rate was 57.3% in 2021 (Raza et al., 2022). Many organisations use the following formula to calculate the rate of attrition:

$$\text{Attrition Rate} = \frac{\textit{Number of employees who left in the year}}{\textit{Average employees in the year}} \times 100. \quad (1)$$

(source: Bhardwaj & Singh, 2016)

During this literature review, employee engagement has been seen having the significant relationship to employee attrition rate, because comparing to engaged employees – 73% of disengaged employees are looking for jobs, while only 37% of engaged employees would consider of new job opportunities. Based on that, it means, that not engaged employees are more likely to change their jobs. Employee engagement might be explained as the matter, when employee enjoy of what they are doing in their work and feel useful and appreciated (Clack, 2020). As five top global drivers of employee engagement according to Phillips et al. (2016) are: management and communication between the company and its employees, the company's reputation, the balance of life and work, and the company's strategy, goals, and objectives. Many companies entrust the responsibility to manage the employee engagement to the first line managers. A leader who inspires, strengthens, and connects employees can help shape employees' perceptions of their work environment as resourceful (Nikolova et al., 2019). According to Clack (2020), engaged employees are more loyal, improvements oriented, keen on to do some extra steps, take challenges, and to speak out about problems. Disengaged employees, on the other hand, require an additional incentive to become high-performing employees. The author emphasizes that employee engagement may be regarded as one of the keys to organisational success. According to Gallup, approximately 70% of employees in the United States are not engaged in their work. This means a significant loss in productivity, with the cost ranging between $450 and 550 billion. Based on this, I must note that this is a big loss even for large organisations. Moreover, employee engagement significantly influences employee performance (Hutama & Sagala, 2019). As a fact – Gallup's survey data shows, that engaged employees are 17% more productive comparing with disengaged (Peters, 2019).

2. Potential solutions used to solve business problems

Over the last years, computers have become more powerful, algorithms are now such advanced and it is possible to connect datasets to enable deeper analyses. It is now possible to collect data from various aspects of the company, ranging from operational workflows to customer behaviour, and to visualize the data using data science principles and data mining techniques (Provost & Fawcett, 2013). Data science, or Artificial Intelligence (AI) is going to solve biggest humanity's challenges. There are three main reasons, why business should use AI: to better understand customers, improve products and services and to apply business process automation (Ward & Marr, 2019). Moreover, AI enables the machines to make predictions, using information learned from historical data (Raza et al., 2022). Every data science problem is different, thus, to solve the specific task, different algorithms should be categorized (Berthold, 2020). A goal of data science, according to Provost and Fawcett (2013) is to improve decision making, which is direct interest of business. Data-driven decision-making (DDD) is when decisions are based on the analysis of data, rather that intuition, therefore, DDD in essence improves business performance.

2.1. Variety of data science techniques and models

Provost and Fawcett (2013) have listed the main analytical techniques:

- Statistics – used to get numeric values of interest from data.
- Database querying – used to get a subset of data or statistics about data, through frontend to database, by creating a query, formulated in technical language.
- Data warehousing – used to collect data across organisation from multiple transaction-processing systems.
- Regression analysis – used to estimate the relationship between variables.
- Machine learning (ML) and data mining – used to analyse data from environment and to learn to predict unknown quantities from that data for making future predictions.

All these techniques – from data collection to ML algorithms might help to discover the model for employee attrition prediction problem. ML algorithms can be used to identify and predict some specific variable, related to the problem (Witten et al., 2011). In data science, prediction generally refers to the process of estimating an unknown value or forecasting a future event; thus, a predictive model may be used to obtain an unknown value of interest in data science (Provost & Fawcett, 2013). Based on this, ML technique is crucial, whereas it is automated process which may consist of different algorithms, used for decision-making, and it provides predictions as the outcome. The goal of ML to get better predictions, than humans (Raza et al., 2022), thus ML has come to play an exceptionally large role in the field over the years (Provost & Fawcett, 2013). Nowadays the business is such highly competitive, customers are high demanding, and economy is service oriented, thus these predictions might really help business to grow (Witten et al., 2011). Data mining and ML algorithms might be well applied in different industries, such as finance, education, healthcare, and IT (Alsheref et al., 2022). There could be diverse types of ML models (Provost & Fawcett, 2013), but since there is no single model, which works perfectly, Berthold (2020) suggests using a grouping for these models. The main models, according to Provost and Fawcett (2013) are: classification/probability estimation, link prediction, regression, similarity matching, profiling, clustering, co-occurrence grouping, data reduction and causal modelling.

Moreover, there are variety of software suites and tools. Some of them are better suited for batch processing and automation, while others are better suited for data analyst support via graphical interface and reporting. Data analysis requires both – software and skills (Berthold, 2020). It is important to analyst to be able to formulate the problems, design experiments, make reasonable assumptions, and to analyse results (Provost & Fawcett, 2013). Therefore, it is important to combine tools and knowledge to perform robust data analysis (Berthold, 2020).

2.2. Earlier studies with applies techniques and tools, used to predict employee attrition

There have been several previous studies on the employee attrition problem, using various models and algorithms. Earlier studies' findings, outcomes, and accuracy were all different. The following is a review of previous studies, including a description of the most used algorithms and techniques for predicting employee attrition:

- Setiawan et al. (2020) performed "HR analytics: Employee attrition analysis using logistic regression" research. They used logistic regression (LR) to identify the factors, which determines job satisfaction and makes impact on engagement. LR is classification algorithm used to describe connection between two variables – dependent and independent (Raza et al., 2022). As the outcome, Setiawan et al. (2020) found out variables (factors), which have impact on employee attrition: overtime, the frequency of business travels, the total number of years spent in the company, the total number of companies worked for, the years with the current manager, low job satisfaction, department, and marital status. According to the authors, employees who are single, have fewer years of employment, and have less experience are more likely to change jobs. Developed model has 75% accuracy, 73% sensitivity and 75% respectively. As the takeaway from the research, the authors recommend HR department to evaluate job environment and satisfaction, employee workload, and interaction between manager and employee to reduce employee attrition rate in the company.
- Another research was performed from a bit distinct perspective. Alsheref et al. (2022) have executed multiple ML algorithms to discover the most effective algorithm to predict employee attrition. Authors used Multilayer Perceptron Classifier (MLP) and Ensemble Classification (EC) models, together with Gradient Boosting (GB) and Random Forest (RF) algorithms. MLP model is artificial neural network model, which introduces how neurons can work. This model has input, production, and concealed layers (Alsheref et al., 2022). EC is used to classify two types of variables into the patterns and match them to compare to variable being analysed (Barvey et al., 2018). GB is used for both regression and classification models, which first trains dataset, and then it estimates some prediction, based on the trained data (Alsheref et al., 2022). The same as GB, RF could be both - classification and regression algorithm. RF is used to create multiple decision trees, which predicts the best answer (Usha & Balaji, 2021). The RF technique yielded the highest accuracy for this study. However, the study's findings revealed that there is no single model that can be considered ideal and perfect for every business context until now. The authors treat their model as optimal but still yet suggested further studies on the topic.
- Usha and Balaji (2021) used classification and clustering models, and different ML algorithms for preparing the prediction model for employee attrition: Naïve Bayes, Decision Tree (DT), J48, K-Means and RF. Naïve Bayes is classification algorithm, based on Bayes Theorem. This algorithm is based on mathematical equation, and typically it states, that each event is independent to any variable (Usha & Balaji, 2021). DT might be both, classification, and regression algorithm, with tree structure. This algorithm learns decision rules from training data, and predicts target class (Raza et al., 2022). J48 is classification, decision tree algorithm which helps to form classification of variables. Using this algorithm, it is possible to identify the highest information gain attribute and classify clear cases. K-Means is clustering algorithm, which gives K clusters as the output. In other words, the K-means algorithm finds k centroids and then assigns each data point to the closest cluster while keeping the centroids with the shortest distance to the centre in mind (Usha & Balaji, 2021). Authors for this study used Python to find correlation of variables. Authors also used Weka (open-source software) to compare the performance of various classification algorithms. They determined that maximum efficiency is achieved by Naïve Bayes algorithm. The authors states, that ML can be identified as a very good tool for developing models for predicting attrition. The study revealed correlation between the variables, related to employee engagement, and found out, that factors such as employee pride, job security, the company's promotion policy, work-life balance, management recognition, and opportunities to growth have the greatest impact on employee attrition rate.
- Fallucchi et al. (2020) performed research, beginning with preparing and analysing the dataset to

be used, then moving on to the design of the prediction model to identify employees who might potentially leave the company. For this study, authors used Naïve Bayes, LR, K-NN, DT, RF, SVM, Linear SVM algorithms. Support vector machine (SVM) is classification algorithm for two-group classification problems. The algorithm chooses support vectors that can be used to create a hyperplane, which serves as a decision boundary, and then finds the best fit for it (Raza et al., 2022). Although the Linear SVM algorithm achieved the highest accuracy (88%), the Naïve Bayes algorithm was determined to be the best classification algorithm, thus 51 of the 71 employees who left the company were correctly predicted. The study uncovered the primary attrition factors such as monthly income, overtime, distance from home to the office and age. Moreover, the authors discovered that employees who are younger, have a lower salary, live a longer distance from work, and have been with the company for less than two years are more likely to change jobs.

The below table contains a summary of the above-mentioned studies (Table 1). In comparison to current

Table 1. Applicability of employee attrition studies to current research

Type of ML	Aim	Target variable	More technique(-s)	Accuracy, %	Independent variables (Employees' information)	Dataset scope	Applicability for the current research	Refs.
LR	To analyse employee attrition using logistic regression and provide findings to understand what should be improved to keep employees in the organisation	Employee attrition	Exploratory data analysis	75%	Demographical and personal data, work conditions, satisfaction, and job role information	4 410	The study's aim matches, similar independent variables, larger dataset scope. Variables, influencing employee attrition identified, conclusions formed. The model's accuracy, however, is medium	Setiawan et al. (2020)
MLP, EC, GB, RF	To test a variety of ML algorithms and present an automated model for predicting employee attrition	Employee attrition	N/A	80–98%, Highest – RF	N/A	1 500	The study's aim partly matches, independent variables not disclosed, similar dataset scope. Variables, influencing employee attrition not identified, conclusions not related to employee attrition. The wide range of ML algorithms tested. The model's accuracy is high	Alsheref et al. (2022)
Naïve Bayes, DT, J48, K-Means, RF	To test a variety of ML algorithms, compare them based on the performance and present an automated model for predicting employee attrition	Plans to continue (work in the organisation)	Data preparation, correlation of variables	Highest – Naïve Bayes 85.98%	Demographical and personal data, work conditions, recognition, satisfaction, job role and work relationship information	N/A	The study's aim partly matches, similar independent variables, not disclosed dataset scope. Variables, influencing employee attrition identified, conclusions formed. Moreover, the wide range of ML algorithms tested. The model's accuracy is high	Usha and Balaji (2021)
Naïve Bayes, LR, K-NN, DT, RF, SVM, Linear SVM	To investigate how objective factors affect employee attrition and to forecast whether a specific employee will leave the company	Employee attrition	Data preparation and cleaning, exploratory data analysis	Highest – Linear SVM – 88%	Demographical and personal data, work conditions, recognition, satisfaction, job role and work relationship information	1 500	The study's aim partly matches, similar independent variables and dataset scope. Variables, influencing employee attrition identified, conclusions formed. Moreover, the wide range of ML algorithms tested. The model's accuracy is high	Fallucchi et al. (2020)

research, Setiawan et al. (2020) research aim had the best match, thus other studies have mostly overlooked automated models based on ML algorithms. During a model development, variety of ML algorithms are typically tested to determine which is the most reliable. Based on ML algorithms tests accuracy from earlier studies, RF, Naïve Bayes and SVM algorithms were identified as reliable algorithms to predict employee attrition. The model never appears only by itself, and there are other techniques for preparing data for model creation or learning more from data, such as data preparation, cleaning, or data exploratory analysis. Most of the other researchers used very similar independent variables to the current study. However, in some of the studies, no other techniques for data preparation and analysis, independent variables or dataset scope were disclosed, raising concerns about the study's representativeness. Moreover, none of the studies mentioned a strategic deployment of the model in business process management, ensuring its continuity.

3. Methodology of employee attrition prediction model development and deployment

The aim of the employee attrition prediction model is to predict potential leavers in the company by training the tool using the dataset with employees' information. To create this employee attrition predictive model, Knime tool was used. Knime is a modular open data science platform, based on visual programming, and its individual tasks are represented by nodes (Berthold, 2020). Knime analytics platform covers majority of ML algorithms: decision trees, random forest, gradient boosted trees, Naïve Bayes, logistic regression, neural networks, SVM etc. (Melcher & Silipo, 2020). For this research, this tool covered all the techniques of model creation, from data preparation and visualization to ML algorithm training, testing, and deployment (Berthold, 2020).

The base and the crucial component of this research is employees' dataset. The open dataset with employees' information was downloaded from Kaggle. Employees' data transformation was done as a part of the preparation for uploading the dataset to Knime. Following that, employees' data exploration analysis was performed in Knime to find the variables that have the greatest impact on employee attrition. Based on earlier studies findings, several different classification and regression algorithms were chosen to determine the most reliable for the current study: Decision Tree Classifier (DTC), Support Vector Machine (SVM), and Naïve Bayes. These algorithms were applied to the current employee attrition prediction model, and their reliability was measured. DTC was founded as the most reliable algorithm, thus it was chosen for future investigation of this research and model deployment.

Model consists of 4 stages: a) data collection, b) data exploration, c) data processing and feature engineering and d) Machine Learning algorithms (model overview is showing in Figure 1).

3.1. Employees' data collection

Dataset with employees' information, used for this research was downloaded from Kaggle (https://www.kaggle.com/code/hamzaben/employee-churn-model-w-strategic-retention-plan/data?select=WA_Fn-UseC_-HR-Employee-Attrition.csv), which is online community platform for data scientists. This dataset size is 227.98 kB. Dataset, provided by IBM HR Analytics Employee Attrition & Performance, consists of 1470 employees. Both sorts of employees are included in the data: those who are currently employed (83.9%) and those who have already quit their jobs (16.1%). Dependent variable in this research is employee attrition. This dataset holds 35 independent variables in total:

– Basic employees' information: identification number, department.
– Employees' demographics: age, distance from

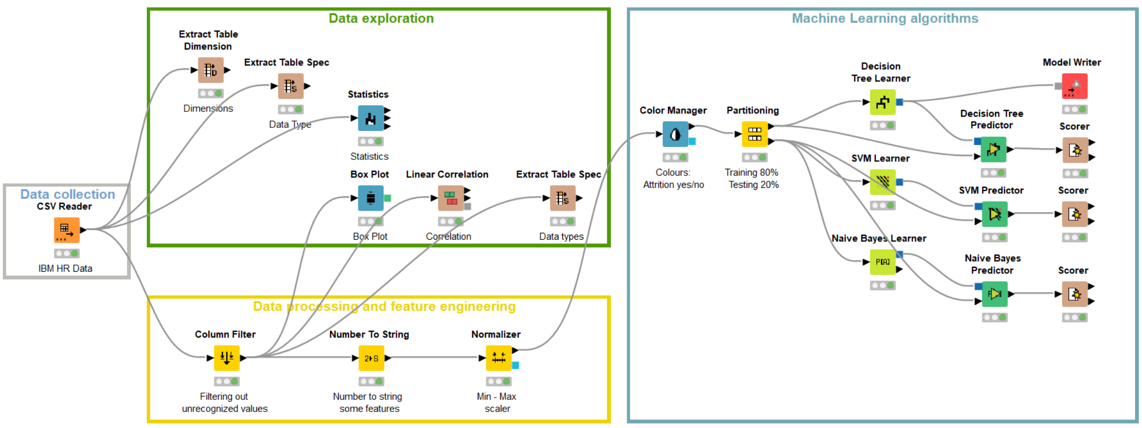

Figure 1. Model overview in Knime

home, education, education field, gender, marital status.
- Employees' satisfaction rates: environment, job, relationship, work-life-balance, and job involvement rates.
- Employees' working conditions: business travel, hourly/daily/monthly rate, monthly income, job level, overtime, percent salary hike, standard hours, stock option level, trainings time last year, years since last promotion.
- Personal employees' experience: job role, performance rating, years at company, total number of working years, years in current role and years with current manager.

First, dataset was downloaded and saved as a comma separated file (csv.). Several independent variables in the original dataset had textual values ('Attrition', 'BusinessTravel', 'Department', 'EducationField', 'Gender', 'Over18', 'MaritalStatus', 'Overtime', 'JobRole'). Therefore, these values were transformed into numeric, from 0 to n – 8. For instance, 'Attrition' variable consisting of 2 values – 'Yes' and 'No'. These values were transformed to '0' (mapped to 'No'), and '1' (mapped to 'Yes'). This was done to be able to build the proper model for data training and testing. Employees' dataset was uploaded to Knime using CSV Reader node afterwards.

3.2. Employees' data processing and feature engineering

During this stage, employees' data processing and feature engineering nodes were used. Column Filter node was used to eliminate unnecessary employees' information from the dataset, such as 'EmployeeCount', 'EmployeeNumber', 'Over18', 'StandHours'. Number to String node converted attrition related variables to strings. Min-max normalization was applied with Normalizer node, with values from Min 0.0 to Max 1.0. This stage of the model is important to prepare the data for employees' data exploration and further model development.

3.3. Employees' data exploration

A variety of summary statistics were reviewed during the employees' data exploration stage to determine the factors that have the greatest impact on employee attrition. Extract Table Dimensions and Extract Table Spec nodes were used to obtain data specifications. The Statistics node was applied, which computes statistical moments across all numeric columns (minimum, maximum, mean, standard deviation, variance, median, overall sum, number of missing values, row count) and counts all nominal values and their occurrences. The following are notable observations made after analysing employees' data through Statistics node output:

- Basic employees' information: employees belong to three departments in total (research and development, sales, and human resources). Most of the leavers are from the research and development department.
- Employees' demographics: the average age of employees who are still working in company is 37.6 years and 33.6 years is the average age of those who have already left the company. The journey for employees from their home to workplace differs from 1 and 29 miles. By fact, most of the leavers lived less than 10 miles from their place of employment. Education fields (life sciences, medical, human resources, marketing) and technical degrees, as well as education levels, are the part of employees' data as well. There are 68% of the leavers who have a higher education level. The data includes 63% males and 37% of females who have already left the company. The dataset, also, includes three marital statuses: single (470 employees), married (673 employees) and divorced (327 employees). At 25%, single employees have the highest proportion of leavers.
- Employees' satisfaction rates: a ranking is associated to all employee's satisfaction/involvement and work-life balance variables, such as 1 'Low', 2 'Medium', 3 'High' and 4 'Very High'. Based on the statistics, employees who are more engaged in their work are less likely to leave the company. Furthermore, as the score for job satisfaction increases, proportion of leavers decreases. Most of the leavers have a 'Low' work-life balance, according to the data.
- Employees' working conditions: a preliminary examination of the relationship between business travel frequency and employee attrition reveals that business travellers accounts for the greatest proportion of the leavers. There is no disclosure of travel frequently or duration metrics, related to business travel, but data shows, that there is 88% of the leavers, who were having business travels. Furthermore, according to the statistics, employees with lower salaries are more likely to leave the company. Moreover, lower percent of salary hike impacts on employee attrition as well.
- Personal employees' experience: Employees are assigned a level within the company ranging from '1' (junior employee) to '5' (managerial employee). Employees with a job level of '1' have the highest proportion of the leavers (62%). Moreover, there is 43% of the employees, who no longer works for the company after working there for less than three

years.

Continuing the statistical analysis, a box plot node was used to display statistical parameters that are insensitive to extreme outliers: minimum, lower quartile, median, upper quartile, and maximum. The correlation of two variables was measured using the Linear Correlation node, by assigning a correlation coefficient to each of the selected columns. Correlation of those variables is showing in Figure 2. The correlation measure produces results that are dependent on the types of underlying numerical variables. Values might vary from –1 to 1. The value of the measure ranges from –1 indicated in red (strong negative correlation) to 1 indicated in blue (strong positive correlation). A value of 0 represents that there is no linear correlation at all (not identified by the colour).

The strongest correlation between employee attrition and other variables is showing in Figure 3. As the outcome of this, it is possible to find the strongest positive correlation of variables to employee attrition such as: overtime, business travelling, education field, distance from home to the office and job role. And the strongest negative correlation of variables to people leaving the company includes total working years, job level, years in current position, monthly income, and years with current manager. A strong negative correlation indicates that there is a strong connection between the two variables, and whenever one rises the other falls. In this case, it means that the aforementioned factors have the greatest impact on employee attrition, which has a strong negative correlation.

The last column in the below table defines *p*-value

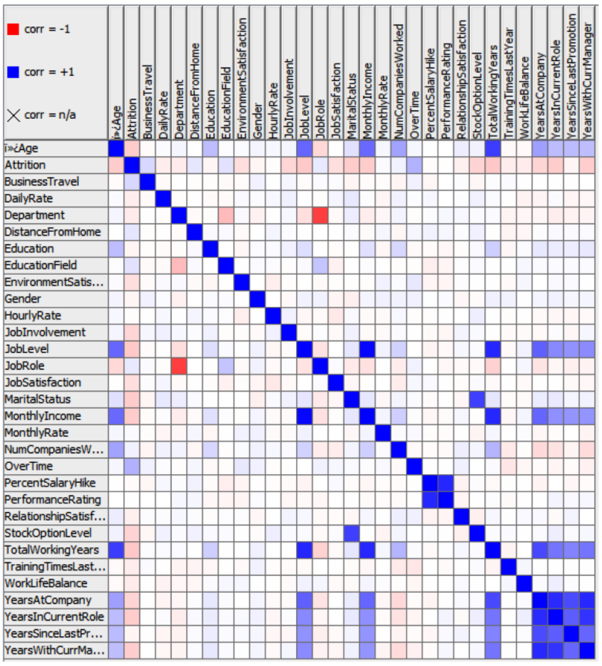

Figure 2. Employee attrition correlation matrix

of correlated variables. *P*-value denotes the likelihood that a statistical measure (e.g.: mean or standard deviation) from an assumed probability distribution will be greater/less than or equal to the observed results. The lower *p*-value means the greater statistical significance of the observed difference. Test hypothesis should be rejected when test result is statistically significant ($P \leq 0.05$). A *p*-value greater than 0.05 indicates that no effect was found.

Row ID	S ▲ Firs...	S Second column name	D ▾ Correlation v...	D p value
Row45	Attrition	OverTime	0.24611799424581...	0.0
Row29	Attrition	BusinessTravel	0.12700648315243...	1.033477590395293E-6
Row34	Attrition	EducationField	0.08961925225095...	5.815959753756594E-4
Row32	Attrition	DistanceFromHome	0.07792358295570...	0.0027930600802130...
Row40	Attrition	JobRole	0.06715149504957...	0.010014034975791786
Row44	Attrition	NumCompaniesWorked	0.04349373905781...	0.09552526205651235
Row43	Attrition	MonthlyRate	0.01517021253047...	0.5611235982242828
Row47	Attrition	PerformanceRating	0.00288875171108...	0.9118840421067675
Row37	Attrition	HourlyRate	-0.0068455495721...	0.7931347689944328
Row46	Attrition	PercentSalaryHike	-0.0134782020574...	0.6056128238894034
Row36	Attrition	Gender	-0.0294532531751...	0.25909236414147463
Row33	Attrition	Education	-0.0313728196400...	0.22931520332230537
Row55	Attrition	YearsSinceLastPromotion	-0.0330187751425...	0.20578995916249176
Row48	Attrition	RelationshipSatisfaction	-0.0458722788811...	0.07871363048462646
Row30	Attrition	DailyRate	-0.0566519918676...	0.029858160660251157
Row51	Attrition	TrainingTimesLastYear	-0.0594777985564...	0.022578499737193107
Row52	Attrition	WorkLifeBalance	-0.0639390472174...	0.014211054989015176
Row31	Attrition	Department	-0.0639905963380...	0.014133018076800091
Row35	Attrition	EnvironmentSatisfaction	-0.1033689783379...	7.172338549366209E-5
Row41	Attrition	JobSatisfaction	-0.1034811260690...	7.043066741729775E-5
Row38	Attrition	JobInvolvement	-0.1300159567860...	5.677065356741631E-7
Row53	Attrition	YearsAtCompany	-0.1343922139899...	2.318871610385602E-7
Row49	Attrition	StockOptionLevel	-0.1371449189333...	1.3010149660019162...
Row56	Attrition	YearsWithCurrManager	-0.1561993159016...	1.7369867845235953...
Row42	Attrition	MonthlyIncome	-0.1598395823849...	7.147363985353159E...
Row54	Attrition	YearsInCurrentRole	-0.1605450042677...	6.003185843639153E...
Row39	Attrition	JobLevel	-0.1691047509310...	6.795384780012408E...
Row50	Attrition	TotalWorkingYears	-0.1710632461362...	4.061878111266275E...

Figure 3. Variables correlation to employee attrition

3.4. Machine Learning algorithms

In the current study, three ML algorithms (Decision Tree, SVM and Naïve Bayes) were applied to identify which is the most reliable to predict employee attrition probability. The selection of ML algorithms is based on earlier studies and the findings acquired. Additional preparation, such as colour management and partitioning were used before the running the algorithm. Color Manager node was used to assign colours for nominal values. The values were computed during the execution process. Red colour was assigned for employee attrition value '1', and green for employee attrition value '0'. Partitioning node was used to split the input table into two partitions to train (80%) and test (20%) data. The two partitions were split and became available at the two output ports.

To induce a classification decision tree in main memory, the Decision Tree Learner node was used. The nominal attribute 'Attrition' was the target attribute. This algorithm's calculations were divided into two quality measures: the gini index and the gain ratio. Furthermore, the post pruning method was used to reduce the tree size and improve prediction accuracy. The pruning method is based on the principle of minimum description length. The Decision Tree Predictor node predicted the class value for new patterns using an existing decision tree. This predictive algorithm is distinct from the others in that it empowers the overview of data in a tree structure, allowing it to clearly see what influences the employee attrition.

The SVM Learner node uses the input data to train a support vector machine. This node manages multi-class problems by computing the hyperplane between each class and the others. SVM Learner node generated predictions for given values and thus provided output to SVM Predictor node about employee attrition probability. The same data columns were used for both training and prediction. The output table includes an extra column for prediction and, optionally, class probabilities.

Naïve Bayes is classification algorithm, based on Bayes Theorem and the below equation:

$$P\left(\frac{y}{x}\right) = \frac{P\left(\frac{x}{y}\right) \times P(y)}{P(x)},$$

(source: Usha & Balaji, 2021) (2)

where: x is the attribute; y is the class; $P\left(\frac{y}{x}\right)$ stands for probability of occurrence of event Y given that event X has occurred (Usha & Balaji, 2021).

Naïve Bayes learner node created a Bayesian model from the given training data. To predict the class membership of unclassified data, the created model was used in the Naïve Bayes predictor. The probability for each attribute and the probability for the class attribute alone make up the class probability. The probability for nominal values is calculated by dividing the total number of instances of the class value by the number of instances of the class value with the specified value. By assuming that each attribute has a normal distribution, the probability of numerical values is estimated.

To test algorithms' reliability, scorer node was used. The confusion matrix is displayed by the scoring node after comparing two columns based on their attribute value pairs. The confusion matrix shows that the classification and row properties are consistent. Hence, the confusion matrix's columns contain the values from the second column, but its rows contain the values from the first column. The output of the node is the confusion matrix, where each cell contains the number of matches. Together with the overall accuracy and Cohen's kappa, the second out-port also provides accuracy statistics for True-Positives, False-Positives, True-Negatives, False-Negatives, Recall, Precision, Sensitivity, Specificity, and F-measure. The summary of algorithms reliability is showing in the below table. Tested ML algorithms are compared in the below Table 2. ML algorithms comparison.

Decision Tree algorithm for employee attrition predictive model was found as the most reliable (97.4% accuracy), then SVM (88.1% accuracy), and least reliable was Naïve Bayes (82.3% accuracy). Model Writer node was the last one task performed under this stage. The purpose of this node is to write a Knime model to a file which can be read with Model Reader node for future model deployment. In this case this node was used only for Decision Tree Learner algorithm, thus this algorithm has been identified as the most reliable.

3.5. Deployment of employee attrition prediction model

As the company collects new data on employed employees and the leavers, the chosen algorithm can be re-trained to generate more accurate predictions, and to identify high-risk employees of leaving the company. This model could be deployed as a tool to predict employee attrition in the company, and a strategic employee attrition plan could be implemented for potential company leavers. To gather required information and to be able to measure employee satisfaction related information, surveys or other tools might be applied. Moreover, other preventive actions should be applied to improve working conditions for high-risk employees. These preventive actions could be considered as a strategic employee attrition plan, with KPIs for employee attrition rate integrated into the PPM framework. To enable this KPI, it is crucial to include management and HR department into this procedure.

Table 2. ML algorithms comparison

Algorithm	Variable	Recall	Precision	Sensitivity	Specificity	F-mean	Accuracy	Cohen's kappa
Decision Tree	Attrition 'Yes'	0.898	0.939	0.898	0.989	0.918	0.974	0.903
	Attrition 'No'	0.989	0.981	0.989	0.898	0.985		
SVM Learner	Attrition 'Yes'	0.34	0.895	0.34	0.992	0.493	0.881	0.44
	Attrition 'No'	0.992	0.88	0.992	0.34	0.933		
Naïve Bayes	Attrition 'Yes'	0.78	0.487	0.78	0.832	0.6	0.823	0.494
	Attrition 'No'	0.832	0.949	0.832	0.78	0.886		

Conclusions

Based on the literature review, PPM is integral part of business process management, as it enables process measurement, monitoring, and control, and it aids in achieving better organisational results by establishing KPIs. Setting KPIs is an appropriate strategic approach for maintaining a consistent strategic approach to achieving a set target. A unified intelligence approach, on the other hand, is required to have real-time status of process performance and to coordinate the organisation's assets. For this reason, tool to help monitor and measure set KPIs should be developed.

Reviewed literature analysis revealed that ML might be perfectly applied to solve employee attrition problem, by using predictive model. The earlier studies of employee attrition problem shows that classification and regression models are the most frequently used ML models for this specific problem.

For this research, dataset, extracted from Kaggle, of 1470 employees was used, to develop employee attrition probability model. Employee's data preparation, transformation, and data exploration analysis were carried out to identify the variables that have the greatest impact on employee attrition. Employee's data exploration analysis of the current dataset revealed that employees, who are single, younger, or with higher education were more likely to leave the company. Moreover, employees who travel for business have a higher proportion of leavers than their colleagues. Furthermore, employee attrition is influenced by work-life balance, as well as job involvement and environmental satisfaction. Besides, employee attrition is influenced by salary and percentage of salary increase. Further, the biggest part of not active employees was holding lower job positions, and most not active employees worked less than 3 years at the company. The strongest positive and negative correlations between employee attrition and other variables were identified following data exploration analysis. According to industry standards, the dataset was divided into two parts: 80% for training and 20% for validation. Based on the earlier employee attrition studies, three different ML algorithms (DTC, SVM and Naïve Bayes), were applied to identify the most reliable algorithm. The most reliable algorithm was DTC, with 97.4% accuracy in predicting employees who might leave the company. Since this is a self-learning algorithm, accuracy is likely to change over the time as more datasets are added. DTC algorithm has been chosen for model deployment, and its purpose is to predict employee attrition using new data.

Based on literature analysis and current research, the following recommendations to reduce employee attrition rate ensuring continuity might be provided to business: first, employee attrition, along with its target value, might be established as an organisational KPI to be integrated into the PPM framework to manage its rate. Second, the developed ML predictive model might be applied, as it is accurate tool to predict employee attrition. The aim of this tool is not only to predict potential leavers, but also to measure and monitor employee attrition KPI. A variety of ML algorithms should be tested before selecting the most reliable ML algorithm. However, it is critical to combine tools and knowledge to develop predictive model and apply statistical analysis. Third, a strategic employee attrition plan might be implemented and applied for identified potential company leavers based on deployed model predictions. Employee, job, and relationship satisfaction measurements, as well as other employee attrition data, should be gathered through surveys or conversations between management or HR representatives and employees. A strategic employee attrition plan should assist in lowering the employee attrition rate and controlling the employee attrition KPI. These recommendations are not intended to be a one-time fix, but they may aid in reducing employee attrition and ensuring that the employee attrition target is consistently met.

References

Alao, D., & Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. *Computing, Information Systems & Development Informatics*, *4*(1), 17–28. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1012.2947&rep=rep1&type=pdf

Al-Darraji, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., & Abdulmalik, H. A. (2021). Employee attrition prediction using deep neural networks. *Computers*, *10*(11), 141. https://doi.org/10.3390/computers10110141

Alsheref, F. K., Fattoh, I. E., & Ead, W. M. (2022). Automated prediction of employee attrition using ensemble model based on machine learning algorithms. *Computational Intelligence and Neuroscience*, *2022*, 7728668. https://doi.org/10.1155/2022/7728668

Balaban, N., Belić, K., & Gudelj, M. (2011). Business process performance management: Theoretical and methodological approach and implementation. *Management Information Systems*, *6*(4), 3–9. https://www.ef.uns.ac.rs/mis/archive-pdf/2011%20-%20No4/MIS2011_4_1.pdf

Barvey, A., Kapila, J., & Pathak, K. (2018). *Proactive intervention to downtrend employee attrition using artificial intelligence techniques*. https://arxiv.org/abs/1807.04081

Berthold, M. R. (2020). *Guide to intelligent data science: How to intelligently make use of real data* (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-45574-3

Bhardwaj, S., & Singh, A. (2016). Factors affecting employee attrition among engineers and non-engineers in manufacturing industry. *Journal B&IT*, *7*(2), 26–34. https://doi.org/10.14311/bit.2017.02.04

Brenes, E. R. (2000). Strategies for globalizing Latin American business. *Journal of Business Research*, *50*(1), 3–7. https://doi.org/10.1016/S0148-2963(98)00102-7

Clack, L. (2020). Employee engagement: Keys to organizational success. In *The Palgrave handbook of workplace well-being*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-02470-3_77-1

Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. Predicting employee attrition using machine learning techniques. *Computers*, *9*(4), 86. https://doi.org/10.3390/computers9040086

Fleischmann, A. (2012). *Subject-oriented business process management*. Springer. https://doi.org/10.1007/978-3-642-32392-8

Folan, P., & Browne, J. (2005). A review of performance measurement: Towards performance management. *Computers in Industry*, *56*(7), 663–680. https://doi.org/10.1016/j.compind.2005.03.001

Gruman, J. A., & Saks, A. M. (2011). Performance management and employee engagement. *Human Resource Management Review*, *21*(2), 123–136. https://doi.org/10.1016/j.hrmr.2010.09.004

Haldorai, K., Kim, W. G., Pillai, S. G., Park, T., & Balasubramanian, K. (2019). Factors affecting hotel employees' attrition and turnover: Application of pull-push-mooring framework. *International Journal of Hospitality Management*, *83*, 46–55. https://doi.org/10.1016/j.ijhm.2019.04.003

Hey, R. B. (2017). What is performance management? In *Performance management for the oil, gas, and process industries: A systems approach* (pp. 1–9). Elsevier. https://doi.org/10.1016/B978-0-12-810446-0.00001-3

Hutama, J. A. N., & Sagala, E. J. (2019). Influence of employee engagement and organizational culture towards employee performance. *Manajemen Bisnis*, *9*(2), 107–113. https://doi.org/10.22219/jmb.v9i2.7939

Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: A machine learning approach. *SN Applied Sciences*, *2*(4), 757. https://doi.org/10.1007/s42452-020-2519-4

Melcher, K., & Silipo, R. (2020). *Codeless deep learning with KNIME: Build, train, and deploy various deep neural network architectures using KNIME analytics platform* (Community ed.). Packt Publishing.

Nikolova, I., Schaufeli, W., & Notelaers, G. (2019). Engaging leader – Engaged employees? A cross-lagged study on employee engagement. *European Management Journal*, *37*(6), 772–783. https://doi.org/10.1016/j.emj.2019.02.004

Ongena, G., & Ravesteyn, P. (2020). Business process management maturity and performance. *Business Process Management Journal*, *26*(1), 132–149. https://doi.org/10.1108/BPMJ-08-2018-0224

Peters, J. (2019). Employee engagement. In *Knowledge resources* (pp. 2–24). KR Publishing. https://web.s.ebscohost.com/ehost/detail/detail?vid=0&sid=4dd97d56-66a4-44d1-9f4a-904176a21d2f%40redis&bdata=JmF1dGh0eXBlPWNvb2tpZSxjcGlkJmN1c3RpZD1zNjgyNDM4NyZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#AN=2322279&db=e000xww

Phillips, J. J., Phillips, P. P., & Ray, R. (2016). *Measuring the success of employee engagement: A step-by-step guide for measuring impact and calculating ROI*. Association for Talent Development. https://web.s.ebscohost.com/ehost/detail/detail?vid=0&sid=d266a4b3-dc7c-4336-90c1-9b379b6d7f6b%40redis&bdata=JmF1dGh0eXBlPWNvb2tpZSxjcGlkJmN1c3RpZD1zNjgyNDM4NyZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#AN=1222510&db=e000xww

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly.

Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. *Applied Sciences*, *12*(13), 6424. https://doi.org/10.3390/app12136424

Setiawan, I., Suprihanto, S., Nugraha, A. C., & Hutahaean, J. (2020). HR analytics: Employee attrition analysis using logistic regression. *IOP Conference Series. Materials Science and Engineering*, *830*(3), 32001. https://doi.org/10.1088/1757-899X/830/3/032001

Usha, P. M., & Balaji, N. V. (2021). A comparative study on machine learning algorithms for employee attrition prediction. *IOP Conference Series. Materials Science and Engineering*, *1085*(1), 12029. https://doi.org/10.1088/1757-899X/1085/1/012029

Van der Aalst, W. M. P., la Rosa, M., & Santoro, F. M. (2016). Business process management: don't forget to improve the process. *Business & Information Systems Engineering*, *58*(1), 1–6. https://doi.org/10.1007/s12599-015-0409-x

Ward, M., & Bernard Marr, M. (2019). *Artificial intelligence in practice: How 50 successful companies used artificial intelligence to solve problems*. Wiley-Blackwell.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Elsevier/Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-374856-0.00001-8

Yahia, N. B., 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. https://doi.org/10.1109/ACCESS.2021.3074559

PROCESŲ VEIKLOS VALDYMO TOBULINIMAS, PROGNOZUOJANT DARBUOTOJŲ KAITĄ TARPTAUTINĖJE ĮMONĖJE

Aleksei IURASOV, Virginija BARGAILĖ

Santrauka. Darbuotojų kaita tapo didele problema organizacijoms, nes daro įtaką jų procesų veiklai, taip pat sukelia finansinius bei verslo valdymo nuostolius. Be to, pastebima tendencija, jog ši problema auga, todėl labai svarbu rasti būdą sumažinti darbuotojų kaitos rodiklius. Vienas iš galimų sprendimų sumažinti šį rodiklį yra prognozuojant darbuotojų kaitą organizacijose ir pritaikant tinkamus sprendimus darbuotojams, galimai norintiems palikti organizaciją. Mašininio mokymosi modelis ir algoritmai gali būti pritaikomi ir naudojami darbuotojų kaitai prognozuoti. Šios prognozės yra daug tikslesnės nei vien prielaidos, grįstos intuicija, nes mašininio mokymosi algoritmai atlieka prognozes, remiantis istoriniais duomenimis. Kadangi egzistuoja skirtingų mašininio mokymosi algoritmų įvairovė, šioms prognozėms turi būti pasirinktas pats patikimiausias algoritmas. Taigi, šiam prognozavimo modeliui Sprendimų medžio algoritmas buvo identifikuotas kaip pats patikimiausias (prognozuojantis 97,4 % tikslumu). Deja, vien turint darbuotojų kaitos prognozavimo modelį dar nereiškia, jog darbuotojų kaitos problema bus išspręsta. Šis modelis kaip įrankis turėtų būti integruotas į organizacijos verslo procesų valdymą, užtikrinant tęstinumą. Procesų veiklos valdymo struktūra gali būti tinkamiausia šiam tęstinumui užtikrinti, nes įgalina procesų matavimą, kontrolę ir padeda pasiekti geresnius rezultatus organizacijoje, nustatant ir matuojant veiklos rodiklius. Veiklos rodikliai yra svarbi procesų veiklos valdymo dalis, nes sujungia verslo veiklą su strategija. Todėl strateginis darbuotojų kaitos prognozavimo sprendimas turi būtinai susisieti su kuriuo nors iš organizacijos veiklos rodiklių, nes taip bus pasiektas tikslas užtikrinant sprendimo tęstinumą.

Reikšminiai žodžiai: proceso veikla, veiklos rodikliai, darbuotojų kaitos prognozavimas, mašininis mokymasis.