VILNIUS GEDIMINAS TECHNICAL UNIVERSITY
FACULTY OF BUSINESS MANAGEMENT
DEPARTMENT OF BUSINESS TECHNOLOGIES AND ENTREPRENEURSHIP

Virginija Bargailė

IMPROVING PROCESS PERFORMANCE MANAGEMENT BY PREDICTING EMPLOYEE ATTRITION IN INTERNATIONAL COMPANY

PROCESŲ VEIKLOS VALDYMO TOBULINIMAS, PROGNOZUOJANT DARBUOTOJŲ KAITĄ TARPTAUTINĖJE ĮMONĖJE

Master's Degree Thesis

Business Management study programme, state code 6211LX058
International Business specialisation
Business study field

Vilnius, 2024

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY
FACULTY OF BUSINESS MANAGEMENT
DEPARTMENT OF BUSINESS TECHNOLOGIES AND ENTREPRENEURSHIP

Virginija Bargailė

IMPROVING PROCESS PERFORMANCE MANAGEMENT BY PREDICTING EMPLOYEE ATTRITION IN INTERNAITONAL COMPANY

PROCESŲ VEIKLOS VALDYMO TOBULINIMAS, PROGNOZUOJANT DARBUOTOJŲ KAITĄ TARPTAUTINĖJE ĮMONĖJE

Master's degree Thesis

Business Management study programme, state code 6211LX058
International Business specialisation
Business study field

Supervisor prof. dr. Aleksei Iurasov ________ ________
(Title, Name, Surname) (Signature) (Date)

Vilnius, 2024

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY
FACULTY OF BUSINESS MANAGEMENT
DEPARTMENT OF BUSINESS TECHNOLOGIES AND ENTREPRENEURSHIP

Business Studies study field
Business Management study programme, state code 6211LX058
International Business specialisation

APPROVED BY
Head of Department
Ieva Meidutė-Kavaliauskienė
2023-12-06

OBJECTIVES FOR MASTER THESIS

No. TVfmu-22–5264

Vilnius

For student Virginija Bargailė

Master Thesis title: Improving Process Performance Management by Predicting Employee Attrition in International Company

Deadline for completion of the final work according to the planned study schedule.

THE OBJECTIVES:

Tasks for the Master's Final Thesis:
To review the existing scientific knowledge on the causes and consequences of employee attrition and the methods and models used to address it.
To design and implement an algorithm that can estimate the value of an employee and the probability of his or her attrition based on various factors and data sources.
To evaluate the performance and accuracy of the algorithm and compare it with other existing models and methods.
To provide actionable recommendations for managers and organizations on how to retain valuable employees and reduce attrition rates.

Academic Supervisor Professor Aleksei Iurasov

Vilnius Gediminas Technical University	ISBN / ISSN
Faculty of Business Management	Copies No. ....
Department of Business Technologies and Entrepreneurship	Date ..........-.....-.....

Master Degree Studies **Business Management** study programme Master Graduation Thesis	
Title	**Improving Process Performance Management by Predicting Employee Attrition in International Company**
Author	**Virginija Bargailė**
Academic supervisor	**Aleksei Iurasov**

	Thesis language: English

Annotation

With this Master Thesis it is analysing employee attrition impact to process performance management in international companies. With the first literature part there is analysing and evaluating the connection and impact of employee attrition to process performance management. Process performance, productivity, and efficiency were all affected due to this influence. With the second literature part there was analysed data science techniques used to solve business problems. Moreover, it was analysed earlier research, applied to employee attrition problem. The analysis and the results of earlier research helped to apply the best practice in the current research and to develop employee attrition prediction model. With the methodology part it was described predictive model development (using an open data set), together with all the development stages, and performed statistical analysis to identify the factors, which impacts employee attrition the most. With the practical-deployment part, additional literature analysis has been performed to develop the strategic deployment of employee attrition prediction model. Employee attrition prediction model has been tested by modifying the data and thus to improving working conditions for potential leaver. The results have been tested by re-calculating employee attrition probability. The results showed zero probability to leave the company, because working conditions have been improved. With the last parts it was provided research limitations, conclusions, and recommendations.

Master Thesis includes 7 parts: introduction, literature analysis, methodology part, practical part, research limitation, conclusions and recommendations and references.

Master Thesis volume – 46 pg. without appendices, 24 figures, 4 tables, 49 references.

In additionally added appendices, scientific article with same topic as this Master Thesis, and certificate of participation in the 26th Conference for Young Researchers “Science – future of Lithuania. Economics and Management”.

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

Vilniaus Gedimino technikos universitetas	ISBN / ISSN
Verslo vadybos fakultetas	Egz. sk. ....
Verslo technologijų ir verslininkystės katedra	Data ..........-.....-.....

Antrosios pakopos studijų **Verslo vadybos** programos magistro baigiamasis darbas	
Pavadinimas	**Procesų veiklos valdymo tobulinimas, prognozuojant darbuotojų kaitą tarptautinėje įmonėje**
Autorius	**Virginija Bargailė**
Vadovas	**Aleksei Iurasov**

Kalba: anglų

Anotacija

Baigiamajame magistro darbe nagrinėjama darbuotojų kaitos įtaka įmonių procesų veiklos valdymui tarptautinėse įmonėse. Pirmoje literatūrinėje dalyje nagrinėjamas ir vertinamas ryšys, tiriama įtaka tarp darbuotojų kaitos ir įmonių procesų veiklos valdymo. Buvo susieta įtaka, kuri pasireiškia proceso našume, produktyvume ir organizacijos procesų efektyvume. Antroji literatūrinė dalis tiria duomenų mokslo technikas ir įrankius, padedančius spręsti verslo problemas. Taip pat analizuojami ankstesni moksliniai tyrimai, taikyti darbuotojų kaitos problemai spręsti. Ištirti ankstesni moksliniai tyrimai ir jų rezultatai padėjo pritaikyti gerąsias praktikas šiame tyrime, ir sukurti darbuotojų kaitos prognozės modelį. Metodologinėje dalyje aprašytas darbuotojų kaitos prognozės modelio kūrimas (naudojant atvirą duomenų šaltinį), detalizuotos visos jo stadijos, atliekama statistinė duomenų analizė, padėjusi suprasti veiksnius, lemiančius darbuotojų kaitą pagal naudojamus duomenis. Praktinėje taikomojoje dalyje buvo atliekama dar viena literatūrinė analizė, kuria remiantis buvo sukurtas ir aprašytas strateginio modelio pritaikymas. Darbuotojų kaitos prognozės modelis buvo ištestuotas, pagerinus darbo sąlygas potencialiai galinčiam palikti įmonę darbuotojui. Modelio rezultatai buvo patikrinti perskaičiuojant prognozė. Buvo nustatyta, jog pagerinus darbo sąlygas, darbuotojo prognozė palikti įmonę išnyko. Paskutinėse darbo dalyse pateikiami tyrimo ribojimai, baigiamojo darbo išvados ir siūlymai.

Darbą sudaro 7 dalys: įvadas, literatūros analizė, metodologinė dalis, praktinė taikomoji dalis, tyrimo ribojimai, išvados ir siūlymai, literatūros sąrašas.

Darbo apimtis – 46 p. teksto be priedų, 24 iliustr., 4 lent., 49 bibliografiniai šaltiniai.

Atskirai pridedami darbo priedai, mokslinis straipsnis šio baigiamojo darbo tema, bei sertifikatas, nurodantis dalyvavimą ir pranešimo skaitymą 26-toje jaunųjų mokslininkų konferencijoje „Mokslas-Lietuvos ateitis. Ekonomika ir Vadyba“.

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

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

	Virginija Bargailė, 20220715	
	(Studento vardas ir pavardė, studento pažymėjimo Nr.)	
	Verslo vadybos fakultetas	
	(Fakultetas)	
	Verslo vadyba, TVfmu-22	
	(Studijų programa, akademinė grupė)	

BAIGIAMOJO DARBO (PROJEKTO)
SĄŽININGUMO DEKLARACIJA

2023 m. gruodžio 06 d.

Patvirtinu, kad mano baigiamasis darbas tema „Procesų veiklos valdymo tobulinimas, prognozuojant darbuotojų kaitą tarptautinėje įmonėje“ yra savarankiškai parašytas. Šiame darbe pateikta medžiaga nėra plagijuota. Tiesiogiai ar netiesiogiai panaudotos kitų šaltinių citatos pažymėtos literatūros nuorodose.

Parenkant ir įvertinant medžiagą bei rengiant baigiamąjį darbą, mane konsultavo mokslininkai ir specialistai: profesorius daktaras Aleksei Iurasov. Mano darbo vadovas profesorius daktaras Aleksei Iurasov.

Kitų asmenų indėlio į parengtą baigiamąjį darbą nėra. Jokių įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs (-usi).

			Virginija Bargailė
	(Parašas)		(Vardas ir pavardė)

CONSENT TO THE USE OF PERSONAL DATA

2023-12-06
(Date)

With this consent I, Virginija Bargailė (hereinafter referred to as the Data subject) agree to the processing of my personal data for the learning purposes of other students by Vilnius Gediminas Technical University, legal entity code 111950243, address – Saulėtekio al. 11, LT-10223 Vilnius (hereinafter referred to as the Data controller). I agree to the processing of the following personal data (mark the relevant item):

☐ name, surname, Bachelor's final thesis;
☐ Bachelor's final thesis without name and surname;
☒ name, surname, Master's final thesis;
☐ Master's final thesis without name and surname.

Personal data processed for this purpose will not be transferred by the Data controller to any third parties, students will be allowed to access the final theses in the internal information system.

The personal data of the Data subject will be used for this purpose for a maximum of 5 years.

By this consent, the Data subject confirms that he / she is aware of the following rights:

- Access to his/her data and how it is processed (right of access);
- Require the rectification or depending on the purposes of the processing of personal data, the addition of incomplete personal data (right of rectification);
- Destroy his/her data or suspend the processing of data (excluding storage) (right to destroy and right to be "forgotten");
- Require the Data controller to restrict the processing of personal data (right to restrict);
- The right to transfer data (right to transfer);
- Do not consent to the processing of personal data when such data are processed or are intended to be processed for other purposes;
- File a complaint with the State Data Protection Inspectorate.

The Data subject has the right to withdraw his consent at any time.

Virginija Bargailė
[Name, surname and signature of Data subject]

Data controller details:
Vilnius Gediminas Technical University
Legal entity code: 111950243
Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania
Tel. (+370 5) 274 5030
E-mail: vgtu@vgtu.lt
VAT payer code: LT119502413
Data protection officer tel. (+370 5) 251 2191, e-mail: dap@vgtu.lt

TABLE OF CONTENTS

TABLE OF CONTENTS 9

LIST OF FIGURES 11

LIST OF TABLES 12

APPENDICES 13

LIST OF ABBREVIATIONS 14

INTRODUCTION 15

1. ANALYSING THE THEORETICAL ASPECTS OF PROCESS PERFORMANCE MANAGEMENT IMPROVEMENT BY PREDICTING EMPLOYEE ATTRITION IN AN INTERNATIONAL COMPANY 17

1.1. The significance of employee attrition in the context of Process Performance Management 17

1.1.1. The importance of Process Performance Management framework. 17

1.1.2. The importance of employee attrition impact on the international company's performance management. 19

1.2. Data science techniques and tools used to predict employee attrition and management interventions to retain employees 22

1.2.1. Possible data science techniques, used to solve business problems. 22

1.2.2. Studies review, with applied techniques and tools, used to predict employee attrition. 25

1.2.3. Earlier studies recommendations and strategic applications to handle with employee attrition. 29

2. METHODOLOGY OF EMPLOYEE ATTRITION PREDICTION MODEL 31

2.1. Model structure 31

2.2. Model development 33

2.2.1. Data processing and feature engineering. 33

2.2.2. Data exploration and visualisation. 35

2.2.3. Machine learning algorithms. 45

3. STRATEGIC DEPLOYMENT OF EMPLOYEE ATTRITION PREDICTION MODEL . 51

3.1. Strategic employee attrition management 51
3.2. Predictive model deployment for strategic employee attrition management 53
RESEARCH LIMITATION 60
CONCLUSIONS AND RECOMMENDATIONS 61
REFERENCES 63
APPENDICES 68

LIST OF FIGURES

Fig. 1. BPM Complexity 17
Fig. 2. Accuracy value for used algorithms. 26
Fig. 3. Distribution of former employees by department. 36
Fig. 4. Distribution of former employees by distance from home to workplace. 37
Fig. 5. Distribution of former employees by education field. 37
Fig. 6. Distribution of former employees by gender. 38
Fig. 7. Distribution of former employees by marital status. 38
Fig. 8. Distribution of former employees by work-life balance score. 39
Fig. 9. Distribution of former employees by business travels frequency. 40
Fig. 10. Distribution of former employees by monthly income and percent salary hike. 41
Fig. 11. Distribution of former employees by employee level. 42
Fig. 12. Distribution of former employees by duration of work. 42
Fig. 13. Employee attrition correlation matrix. 43
Fig. 14. Variables correlation to employee attrition. 44
Fig. 15. Decision Tree view for overtime. 46
Fig. 16. SVM Predictor output table. 47
Fig. 17. Gaussian distribution for BusinessTavel variable per class. 48
Fig. 18. Employee attrition Machine Learning (ML) predictive model 50
Fig. 19. Strategic accountability approach. 51
Fig. 20. Employee attrition prediction model deployment. 54
Fig. 21. Predictions for new applied employees' data. 55
Fig. 22. Identification of potential leavers. 55
Fig. 23. Employee attrition cost estimation tool. 57
Fig. 24. Employee attrition cost comparison. 58

LIST OF TABLES

Table 1. Employee turnover regional statistics in US 20
Table 2. Applicability of employee attrition studies to current research............ 28
Table 3. ML tools ratings breakdown............ 32
Table 4. ML algorithms comparison. 49

APPENDICES

Appendix 1. Dataset values description 68
Appendix 2. Certificate of participation in the 26th Conference for Young Researchers "Science – future of Lithuania. Economics and Management" 72
Appendix 3. Scientific article "Improving process performance by predicting employee attrition in international company" 73

LIST OF ABBREVIATIONS

Artificial Intelligence – AI

Business Process Management – BPM

Business Process Reengineering - BPR

Data-driven decision-making – DDD

Decision Tree Classifier - DTC

Employee Exploratory Data – EDD

Extra Trees Classifier - ETC

Gradient Boosting - GB

Human Resource – HR

K-nearest neighbours - K-NN

Key Performance Indicator – KPI

Logistic regression - LR

Machine learning – ML

Multilayer Perceptron Classifier – MLP

Process Performance Management – PPM

Random Forest – RF

Return on investment - ROI

Support vector machine – SVM

INTRODUCTION

Business environment becomes more competitive (Peters, 2019) therefore to maintain a competitive advantage it is important to retain professional employees (Brenes E. R., 2000; The people make the firm, 2022). However, employees are leaving their jobs for different reasons and according to Jain et al. (2020), it is a relevant problem, because it causes loses of experienced employees. Leavers takes away "know-how", which is the main business advantage, thus knowledge and experience-based organisations should minimize employee attrition to become advantaged against their competitors (Alao & Adeyemo, 2013; Haldorai et al. 2019). Moreover, clients become more demanding, and organisations must obtain the best performance from their employees to meet their productivity, customer experience, quality, and profitability targets (Peters, 2019). Organisations, which are keen on to be high performing, should be innovative, excel at KIPs, should keep key employees, and attract new experienced employees (Phillips J. J. et al., 2016; Haldorai et al. 2019). This problem might be solved or reduced, by predicting employee attrition (Al-Darraji et al., 2021), using data science and big data analytics. Machine learning (ML) algorithms provide predictions based on data, making them more accurate than intuition-based predictions. As a result, ML might be perfectly applied to help to solve the employee attrition problem (Provost & Fawcett, 2013). This solution would assist businesses and their HR managers in taking influencing factors into account and changing approach how they attract new and retain existing employees (Yahia et al., 2021).

Research problem. Employee turnover is increasing, which has a detrimental impact on organisations (Raza et al., 2022). High employee attrition has an impact on productivity and planning continuity (Yahia et al., 2021), because the organisation may lose an efficient employee. Moreover, recruiting new employees necessitates the involvement of human resources (HR), as well as training and integration into the new work environment (Al-Darraji et al., 2021).

Research object. Employee attrition probability identification, and preventive actions which may help to reduce it.

The aim of the thesis. To suggest decision support model, which would help to predict employee attrition and provide the recommendations for organisations on how to reduce employee attrition rate.

Research tasks:

1. Perform literature analysis to identify employee attrition impact to organisation's process performance.

2. Perform literature analysis of decision support models, developed to solve employee attrition problem in earlier studies to understand what methods could be applicable in the current research.
3. Develop reliable predictive model that can forecast employee attrition. For model deployment, the algorithm with the best accuracy should be used.
4. Apply statistical analysis to provide insights about factors, influencing employee attrition.
5. Provide the strategical recommendations on company behaviour, which may help to reduce employee attrition ensuring continuity.

Research Methods.

- A narrative literature analysis was performed to evaluate and interpret literary materials relevant to the influence of employee attrition on process performance. To get comprehensive overview of the available literature on this topic, keyword searches, snowballing technique, and historical analysis were used.
- Systematic literature analysis was performed to identify, evaluate, and synthesize all relevant studies on a research problem. This was accomplished using library catalogues, global search, and article review.
- Statistical analysis was performed using an open data set to acquire a better knowledge of employee data and discover variables that influence employee turnover.
- To develop employee attrition predictive model, quantitative research was performed using ML and the decision tree algorithm.

1. ANALYSING THE THEORETICAL ASPECTS OF PROCESS PERFORMANCE MANAGEMENT IMPROVEMENT BY PREDICTING EMPLOYEE ATTRITION IN AN INTERNATIONAL COMPANY

1.1. The significance of employee attrition in the context of Process Performance Management

1.1.1. The importance of Process Performance Management framework.

Overall interest in process management increased since the 1980s (Harmon, 2014) when global trades expanded, forcing businesses to grow and develop new processes. Business management concepts such as Business Process Reengineering (BPR), Business Process Management (BPM) and Process Performance Management (PPM) were developed in parallel and evolved over the time (Hey, 2017). As a definition, 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 organisation's needs (Hey, 2017). According to Balaban et al. (2011), PPM is a part of BPM, with a complex structure, as it consists of several stages: Process monitoring, Performance measurement, Performance enhancement, Process modelling, Process implementation and Performance planning (Fig. 1 BPM Complexity).

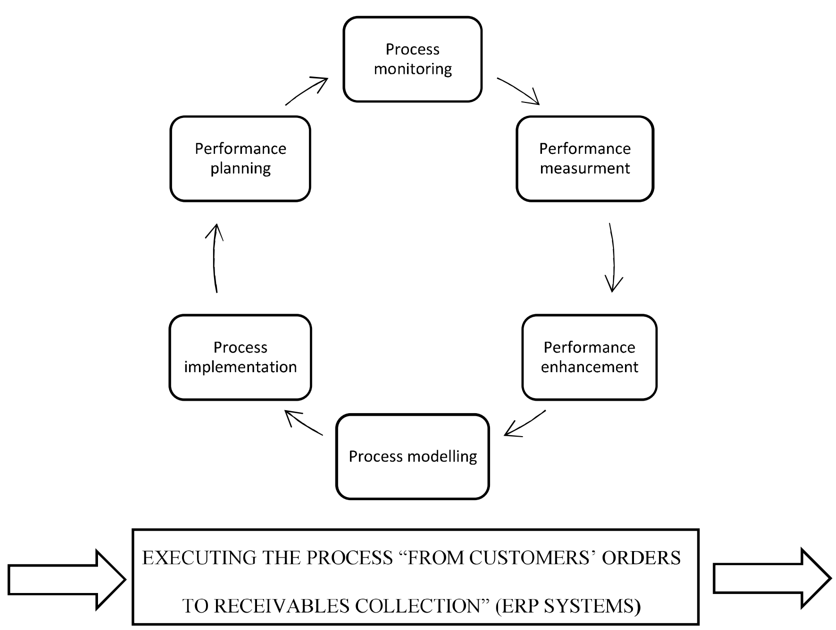


Fig. 1. BPM Complexity

Source: (Balaban et al., 2011)

According to Gruman & Saks (2011), PPM is the concept, through which work is accomplished, thus it is crucial for company's efficiency. This is based on the fact, that PPM enables monitor and control of the process, which is inseparable part to manage business process performance properly (Balaban et al., 2011). However, to be able to monitor the process, appropriate performance measurement and assessment are required. As a result, Key Performance Indicators (KPIs) should be identified and established in the company (Hey, 2017). Balaban et al. (2011) highlights, that process monitoring is exactly the core of BPM cycle because it observes process constantly to achieve planned targets of KPI's, by using surveys, tracking an organisation's activities and processes. Having appropriate performance measurement, 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. Overall, this explains why performance planning, monitoring, measuring, and enhancing is the most important composition of performance control (Balaban et al., 2011).

According to Van der Aalst et al. (2016), KPIs are possible all over the different performance angles such as time, quality, cost, flexibility, etc. According to Fleischmann (2012), it is frequently difficult to develop realistic KPIs, thus the author proposes consultations and experience sharing across organisations before establishing new target values. When establishing new target values, Folan & Browne (2005) proposes involving all organisation's employees - from regular employees to top management. The target values might be established during a series of workshops, brainstorming sessions, interviews, and even pilot process models (Folan & Browne, 2005). The primary aspect, however, is that KPIs would be linked to the organisation's strategy and therefore would benefit the organisation (Fleischmann, 2012). Complexity by setting up KPIs might differ, according to availability of information (e.g., customer satisfaction may require aggregating data from customer experience surveys, etc.). However, unified intelligence approach is necessary, to be able to have real-time status of process performance and to coordinate organisation's assets, related to BPM (e.g.: business processes, processes related data and employee's information) (Balaban et al., 2011).

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 (Folan & Browne, 2005). 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).

To summarize, PPM is important for any organisation since it aids in the improvement of efficiency, quality, customer satisfaction, continuous improvement, compliance and risk management, strategic alignment, and competitive advantage. By measuring, monitoring, and optimizing processes it is feasible to eliminate bottlenecks, decrease waste, and increase production. Efficiency result in faster service delivery which leads to higher customer satisfaction. Furthermore, PPM enables the monitoring of process outputs to ensure quality requirements. To enable measurement and monitoring, organisational KPIs must be established. To have real-time status of process performance and to coordinate the organisation's assets, a unified intelligence approach is also required. Therefore, data, as a background, should drive process performance and provide insights that assist organisations in making well-educated decisions. Proper process management connects processes to the strategic goals of the organisation and guarantees that all organisational decisions contribute to the overarching goals of the business.

1.1.2. The importance of employee attrition impact on the international company's performance management.

Nowadays, the business environment is becoming such competitive (Peters, 2019), that it has the potential to become a major concern in both national and global industries (Brenes E. R., 2000). Because of that, it is important, that companies would realize every available resource when competing with international companies. Employee attrition is an important concern for international businesses because of the financial, operational, and strategic implications. Managing attrition and executing workforce retention and development initiatives are critical for long-term success in the global market. Therefore, it is critical to retain professional employees to maintain a competitive advantage (Brenes E. R., 2000). In the marketing mix (which is a collection of controllable variables, that organisation can use to make an influence on customer), people is one of the most important components (Jain Kumar M., 2013, Išoraitė M., 2021). Based on Išoraitė M., (2021) employees are critical component of the 7Ps marketing mix because they oversee attracting new clients. Furthermore, employees have a positive impact on current client satisfaction with service. In other words, employees have an impact on a company's ability to attract, motivate, and retain customers (Simon et al., 2009).

According to Peters (2019), clients become such demanding, that organisations must obtain the best performance from their employees to deliver the best quality (Peters, 2019). However, the more experienced an employee is, the higher the quality of the service is achieved. Therefore, employees who have competence in their areas are more likely to provide high-quality work (Chyi Lee et al., 2001). Experience is based on each employee knowledge and therefore it plays a strong role (Pousa et al., 2020). As the result, employee's loyalty, and attitude to their work has a strong impact on product quality (Lindsey Hall et al., 2016).

Another critical aspect of business competitiveness is retaining the best performers within the organisation. According to Phillips J. J. et al. (2016), organisations, which are keen on to be high performing, should be more innovative and excel at KIPs (Phillips J. J. et al., 2016). Based on earlier literature, product quality is related to productivity itself. Hence, higher productivity means increased quality. In this context, productivity is defined as the quantity and quality of accomplished work, and employee's productivity is assessed by the number of units of a product or service handled during a given period (Awwad & Heyari, 2022). So, to say, employee attrition affects organisational performance itself (Muhsin & Budaya, 2019), because service may be slowed slightly when employees are leaving, and new employees may take time to get up to speed (Kashif et al., 2017). Company's performance is an indicator of future revenue growth, and therefore productivity has a significant impact on corporate success (Muhsin & Budaya, 2019). Moreover, when employee is productive, this makes impact and increases customer satisfaction as well (Awwad & Heyari, 2022).

Organisations should also keep key employees and to attract new employees by strengthen employee value propositions (Phillips J. J. et al., 2016). However, based on the statistics, it is uneasy to keep loyal employees in the organisation because employee attrition rate in the year 2021 was 57,3% (Raza et al., 2022). Recruiting new employee requires human resource (HR) effort, training, and integration in the new environment (Al-Darraji et al., 2021), moreover the average cost per new hire was \$4129 (Raza et al., 2022). The most common formula to calculate the rate of attrition used by many organisations is (Source: Bhardwaj et al., (2016)):

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

There was significant growth of employee attrition rate observed in 2020, taking 5 years data:

Table 1. Employee turnover regional statistics in US

Region	2016	2017	2018	2019	2020
Northeast	37	38	37	37	53
South	44	46	48	48	57
Midwest	43	42	45	44	58
West	42	42	43	45	58

Source: (27 US Employee Turnover Statistics [2022])

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 top global drivers of employee engagement according to Phillips J. J. 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. Because 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. Moreover, 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).

In conclusion, the relationship between performance management and employees is mutual. Employees who are engaged, informed, and empowered can have a beneficial impact on process performance. Employee involvement, skills, and motivation all have an impact on the quality of products and services. While vice-versa efficient process management can improve employee satisfaction and help the company achieve its goals. The likelihood that an organisation will produce products that meet or exceed customer expectations increases when the well-being, development, and engagement of its employees are prioritized. Ignoring employee-related aspects might result in poor quality, lower customer satisfaction rates, and become potential reputational harm. So, to sum up, it's critical to find a balance that aligns employee interests with PPM goals. However, statistics shows a tendency of growing employee attrition in the organisations, which causes both-financial and business management losses.

1.2. Data science techniques and tools used to predict employee attrition and management interventions to retain employees

1.2.1. Possible data science techniques, 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, M. & Marr, B. (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 & 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. According to Provost & Fawcett (2013) the below there are dominating 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.

A more comprehensive description of these techniques is provided below to determine which of them may possibly solve the employee attrition problem.

Statistics. According to Berthold (2020), statistics is the technique, where data is used for the purpose, to gain new knowledge. To perform statistics, data should be fetched, analysed, interpretated or explained, and presented. The statistical concepts and procedures are divided in the two main areas: descriptive statistics and inferential statistics. The purpose of descriptive statistics is to summarize data without assumptions, making visualisation. While inferential statistics provides more rigorous methods, and the goal of inferential statistics is usually to prepare and support decision making. However, conclusions are only valid if assumptions are satisfied (Berthold, 2020).

Database querying. Database querying purpose is to perform structured search, which is technology for querying multiple data sources independently. Structure of data sources must be identified, and users can formulate queries, control the output order, and access information in real time (Gilula, 2016).

Data warehousing. Data warehouses are typically used to support queries for analysis purposes. It is a database, which considered as historical or real time data of an organisation (Bhatia, 2019).

Regression analysis. Regression technique is used to explain the correlation between analytical variables (Setiawan et al., 2020). According to Witten et al. (2011) it could be two types of regression analysis – linear and logistic. Linear regression can easily be used for classification in domains with numeric attributes. Logistic regression attempts to estimate accurate probability (Witten et al., 2011).

Machine learning (ML) and data mining. Data mining is solving problems by analysing data already presented in databases. Witten et al., 2011 defines it as process of discovering patterns in data, but the patterns must be meaningful, and the process must be automatic or at least semi-automatic (Witten et al., 2011). The ML is fully automated and used together with different algorithms to support decision-making (Raza et al., 2022).

All these techniques – from data collection to ML algorithms might help to develop 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), therefore 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 different 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 below are described different types of data mining models, according to Provost & Fawcett (2013):

- Classification/probability estimation – a model, closely related to scoring, used to determine the class, which individual belongs to.
- Regression – a model, which estimates the numerical value of the variable specific to individual.
- Similarity matching – a model, which identifies similar individuals based on the data.
- Clustering – a model, which based on the data groups individuals by their similarities.
- Co-occurrence grouping – a model, used to find associations between entities based on transactions involving them.
- Profiling – a model, attempts to characterize the typical behaviour or individual, group or population.
- Link prediction – a model, used to predict connections between data items.
- Data reduction – a model, which handles with datasets, replacing big data sets to smaller ones, with much of important information.
- Causal modelling – a model, attempts to help us understand what evens or actions influence others.

Moreover, there are variety of software suites and tools. Some of them are better suited for batch processing and automation, while the others are better suited for data analyst support via graphical interface and reporting. It is important to emphasize that processing with data requires both - software and technical competencies (Berthold, 2020). Analysts must be able analyse data, formulate issues, develop experiments, and make reasonable assumptions (Provost & Fawcett, 2013). Therefore, it is important to combine tools and knowledge to perform robust data analysis (Berthold, 2020). Furthermore, it is critical that models be easily understood and explained, as this can be valuable for conveying results with stakeholders that are unfamiliar with data mining (Provost & Fawcett, 2013).

To summarize all the findings, since AI is used to understand customers, improve products and services, and automate business processes, it may be ideal for solving business problems. AI can tackle business problems by leveraging the power of data, automation, prediction, personalization, and efficiency, ultimately assisting organisations in making better decisions and achieving better results. Overall, statistics, database querying, data warehousing, regression analysis, ML and data mining are the most used data science techniques. Because no single technique is ideal, it is better to employ a combination of these techniques to get the best possible outcomes. Moreover, the combination of software, techniques, and background knowledge is critical. Currently, ML and data mining are regarded as the most effective and active research topics. To handle the employee turnover problem, a variety of techniques and ML algorithms could be applied, ranging from data preparation and analysis to predictive modelling.

1.2.2. Studies review, with applied 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 earlier 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. The model they developed has 75% accuracy, 73% sensitivity and 75% respectively (Raza et al., 2022).
- 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). Both GB and RF could be either classification or regression algorithms. RF is used to create multiple decision trees, which predicts the best answer (Usha & Balaji, 2021). The accuracy, using these algorithms in the research is showing in the below Figure 2. Accuracy value for used algorithms. For this study, the RF

technique yielded the highest accuracy. 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.

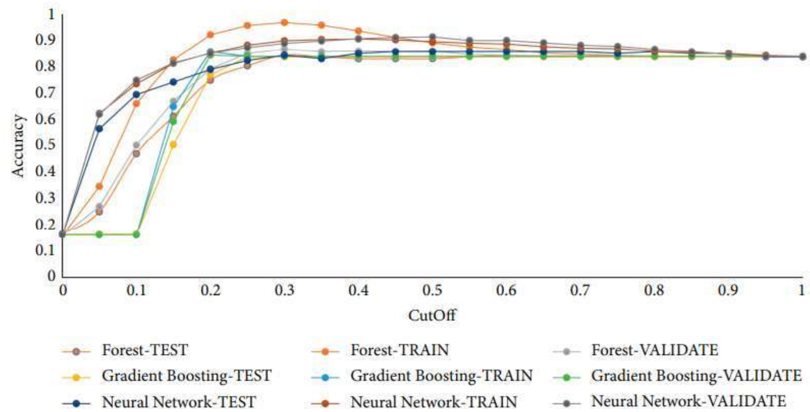


Fig. 2. Accuracy value for used algorithms.

Source: Alsheref et al. (2022)

- Usha & 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). Tree induction is very popular data mining method because it is inexpensive, easy to understand and implement (Provost & Fawcett, 2013). 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.

- Fallucchi et al. (2020) performed research, starting 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, by reason of 51 of the 71 employees who left the company were correctly predicted.

The below table contains a summary of the abovementioned studies (Table 2. Applicability of employee attrition studies to current research). In comparison to current research, Setiawan et al. (2020) research aim had the best match, considering that other studies have mostly overlooked automated models based on ML algorithms. During a model development, variety of ML algorithms are typically tested. From this variety of models and algorithms, used in studies it is possible to frankly presume, that the best model and algorithm should be chosen depending on its reliability and collected data set. So first, variety of models and algorithms should be tested, and the most reliable model and algorithm should be chosen afterwards. 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. Some recommendations on how to lower the employee attrition in an organisation were provided, however, none of the studies mentioned a strategic deployment of the model in business process management, ensuring its continuity.

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 & 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)

1.2.3. Earlier studies recommendations and strategic applications to handle with employee attrition.

Abovementioned researchers identified the reasons of what impacts employee attrition. In all the studies they have found out the variables, which impacts employee attrition the most. Moreover, in some of the research, authors even provided the recommendations on an organisation behaviour to lower employee attrition rate.

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. 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, performed by Usha & Balaji (2021) 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) 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.

Furthermore, some more recommendations and observations have been provided in the others researched, not mentioned in this research before. A model, proposed by Ray A. N. (2019) enabled not only the employee prediction of attrition but also helped to identify the bottlenecks from the analysed dataset. The research disclosed, that employees with highest education level show around 67% of attrition rate in the age group 55-60 years. Relatively high attrition, 38% in the age group 18-24 years. Employees at lower education levels show high attrition levels 48%-50% in the age group of 18-24 years. Alternately, the lowest attrition statistics are observed in the age group of 37-42 years with attrition levels of employees less than or just above 10%. The author also concludes, that having appropriate measures, it is possible to find out the root causes behind the attrition in these employees' groups. Based on that, HR could get some insights and adapt specific actions to lower employee attrition rate.

El-Rayes et al. (2020) during their research have raised some hypotheses, and provided the results, that employee will decide to leave the organisation if offer will be greater than 40% of salary increase. Authors also revealed that such factors as employee engagement and absence are known to have a strong connection with employee attrition (El-Rayes et al., 2020). Moreover, El-Rayes et al. (2020) revealed, that independently to data, tooling, or models, it is very important to offer recommendations to HR professionals based on the studies.

In the last research on employee attrition, performed by Raza et al. (2022) the following factors, which determines employee attrition the most were discovered: when employee get lower salary, is in the age of between 10 to 25 years, working below 1 year at the company and when works higher working hours.

Based on the earlier research findings, it is possible to identify the factors impacting employee attrition the most. Salary, work-life balance, and job satisfaction were listed as the top common factors influencing employee attrition. However, none of the previous employee attrition models were developed along with strategical implementation. Taking all the above-mentioned results and recommendations from employee attrition models, the following actions should be applied to reduce employee attrition rate in the organisation. First, statistical analysis should be performed, that variables, which impacts on employee attrition must be identified. Then, the model which can predict employee attrition should be applied based on different predictive ML algorithms. And finally, the recommendations on organisation behaviour which may help to reduce employee attrition rate should be provided. The solution model must ensure the continuity in this process, therefore it would be applied not a one-time task, but incorporated in to the organisation's strategy.

2. METHODOLOGY OF EMPLOYEE ATTRITION PREDICTION MODEL

2.1. Model structure

Introduction. Since employee turnover can be reduced by being predicted, ML techniques and algorithms should be used to create a predictive model (Provost & Fawcett, 2013; Alsheref et al., 2022). According to Berthold (2020), when constructing an ML model, a combination of data preparation and modelling techniques may be used for the best results. Since there exist variety of different ML tools and algorithms used for predictive models, they should be chosen accordingly (Berthold, 2020). The ML algorithm should be chosen based on its reliability, and the tool should be chosen based on its requirements. The tool should be able to create models using all the methods used in this research, including data preparation, visualization, and training, testing, and deployment of ML algorithms (Berthold, 2020).

The aim of the model: 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.

Objectives of the model.

1. Collect the employee dataset, which consists of current and past employees' information. Variables must include basic employee information, demographics, employee satisfaction, workplace conditions and employees experience information.
2. Process the data and feature engineering to prepare the dataset.
3. Perform data exploration to identify the key factors and trends that impacts on employee attrition.
4. Elaborate the dataset for the training and testing phase and try different classification and regression algorithms to process it.
5. Based on the results collected with test data, compare machine learning algorithm's reliability, and select the algorithm, which fits the best and gives the most accurate results for employee attrition problem.

Tool. The choice of the tool for this research was based on user's evaluation in Gartner's site. Nowadays, the market's three main players who offers predictive analytics tools are: Alteryx Designer (overall rating 4,6/5), KNIME Analytics Platform (overall rating 4,6/5), and Matlab (overall rating 4,4/5). KNIME Analytics Platform has the highest ratings in pricing flexibility and ease of deployment (Source: Gartner, 2023 May). The below table presents ratings of the most important aspects of ML tools (not all ratings are presented) (Table 3. ML tools ratings breakdown).

Table 3. ML tools ratings breakdown

Characteristics	**Alteryx Designer**	**KNIME Analytics Platform**	**Matlab**
Evaluation & Contracting			
Pricing Flexibility	3.8	**4.6**	4.1
Ability to understand needs	4.5	4.4	4.4
Integration & Deployment			
Ease of Deployment	4.4	**4.6**	4.5
Quality of End-User Training	4.5	4.3	4.4
Ease of integration using standard APIs and Tools	4.3	4.3	4.3
Availability of 3rd-Party Resources	4.3	4.3	4.2
Service & Support			
Timeliness of vendor response	4.6	**4.5**	4.4
Quality of Technical support	4.5	4.4	4.5
Quality of Peer user Community	4.7	4.4	4.6

Source: Gartner, 2023 May

Depending on ease and capabilities of model implementation, KNIME (Konstanz Information Miner) tool was chosen for this research. 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).

Background of knowledge. Model has been developed and explained using "Codeless Deep Learning with KNIME" book, written by Melcher, K. & Silipo, R. (2020), and guides in KNIME analytic tool itself.

Model overview. For the best outcomes, the model will include different stages for data gathering, data exploration, data processing, algorithm assignment, and reliability testing.

Algorithms. Based on earlier studies findings, several different classification and regression algorithms is chosen to determine the most reliable for the current research: Decision Tree Classifier (DTC), Support Vector Machine (SVM), and Naïve Bayes.

To summarize the model structure, the KNIME tool was chosen for this research. The DTC, SVM, and Naïve Bayes algorithms were chosen as the most solid for the current investigation based on the results of past studies. The employee attrition prediction model will be used with these algorithms, and the reliability of each will be evaluated to determine which is the most trustworthy. For model deployment, the most reliable algorithm will be chosen.

2.2. Model development

2.2.1. Data processing and feature engineering.

Employees' dataset was first obtained and reviewed because it serves as the foundation and a key component of this research. Dataset with employees' information, used for this research was downloaded from Kaggle as a comma separated file (csv.) (https://www.kaggle.com/code/hamzaben/employee-churn-model-w-strategic-retention-plan/data?select=WA_Fn-UseC_-HR-Employee-Attrition.csv). Kaggle is online community platform for data scientists. The size of dataset is 227.98 kB. Dataset, provided by IBM HR Analytics Employee Attrition & Performance, consists of 1470 employees. Both statuses 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 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.

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 ML model for data training and testing. Data transformation was not necessary for data visualisation. Extensive dataset values description might be found in Appendix.1. Dataset values description. Employees' dataset was uploaded to KNIME using CSV Reader node afterwards.

Data processing and feature engineering stage of the model is important to prepare the data for employees' data exploration and further model development. During this stage, employees' data processing and feature engineering nodes were used. There was used four different nodes: Column Filter, Number To String and Normalizer. Column Filter node was used to eliminate unnecessary employees' information from the dataset, such as 'EmployeeCount', '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.

To sum up data processing and feature engineering stage, the most important component of this stage was dataset. For this research an open dataset, extracted from Kaggle, of 1470 employees was used, to develop employee attrition probability model. Since most of the data variables are sensitive and confidential, using real company data would be restricted. The applied dataset contains 35 independent variables in total, including basic information, demographics, employee satisfaction rates, working conditions, and personal experience. It is critical to prepare the data for employees' data exploration and subsequent model creation during the model's processing and feature engineering stages. Some of the data variables were converted to numbers.

2.2.2. Data exploration and visualisation.

Data exploration and data visualisation stage is important to get comprehensive and consistent data analysis. These stages help to get know the data from current dataset, and to aggregate the data identifying the variables, which are the most related to employee attrition. A variety of summary statistics were reviewed during these stages to determine the factors that have the greatest impact on employee attrition. For data exploration stage it was used five different nodes: Extract Table Dimension, Extract Table Spec, Statistics, Box Plot and Linear Correlation. Extract Table Dimensions and Extract Table Spec nodes were used to obtain data specifications. Extract Table Spec - Node extracting the meta information from the input table (column names, types, etc.). 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.

There following nodes were used for data visualisation: Row Filter, Color Management, Interactive Pie chart (local), Histogram (local) and Scatter Matrix (local). Data visualization is an important research tool since it allows for easier data exploration, analysis, and communication. It helps to obtain insights, make evidence-based decisions, and visualise findings. Employees, who already left the company were filtered out using Row Filter node. Other nodes were used to visualise the data in different ways. Observing the biggest deviations in employee data which are the factors, impacting on employee attrition, they are visualised in the figures below (which are the outputs from data visualisation nodes). The following are notable observations made after analysing employees' data:

- 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 (showing in Fig. 3. Distribution of former employees by department).

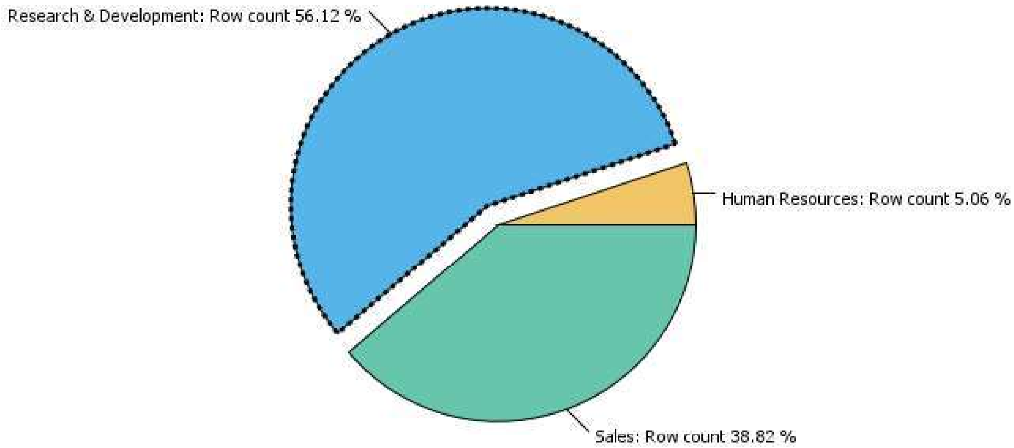


Fig. 3. Distribution of former employees by department.

- 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 workplace (showing in Fig. 4. Distribution of former employees by distance from home to workplace). 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. Employees distribution by education fields is showing in Fig. 5. Distribution of former employees by education field. Most of the leavers had higher education in life sciences field. The data includes 63% males and 37% of females who have already left the company (showing in Fig. 6. Distribution of former employees by gender). 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 (showing in Fig. 7. Distribution of former employees by marital status).

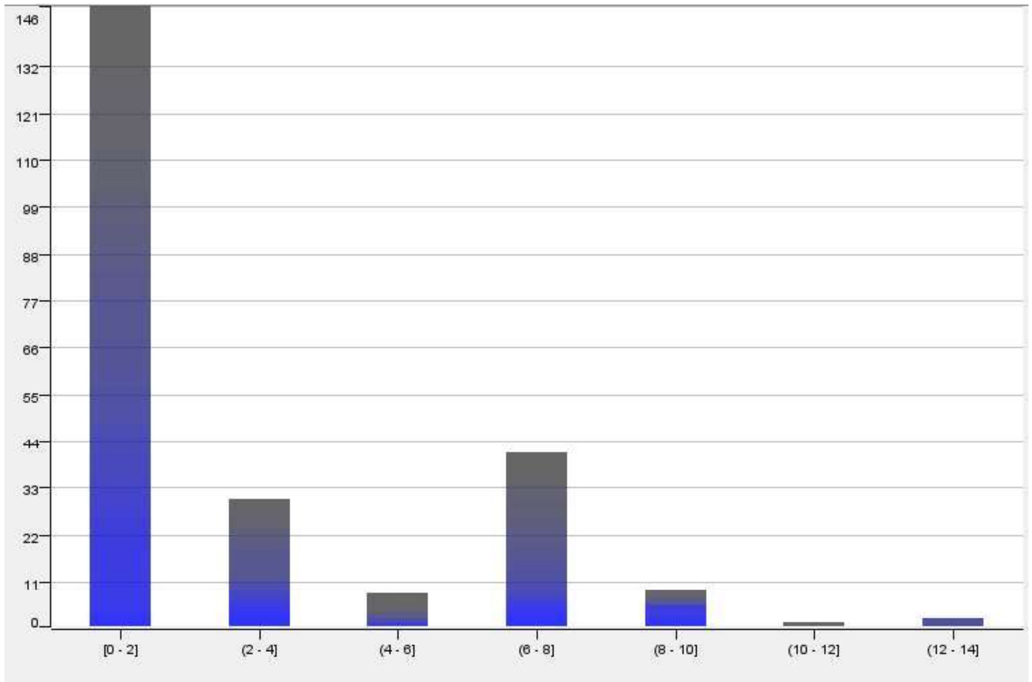


Fig. 4. Distribution of former employees by distance from home to workplace.

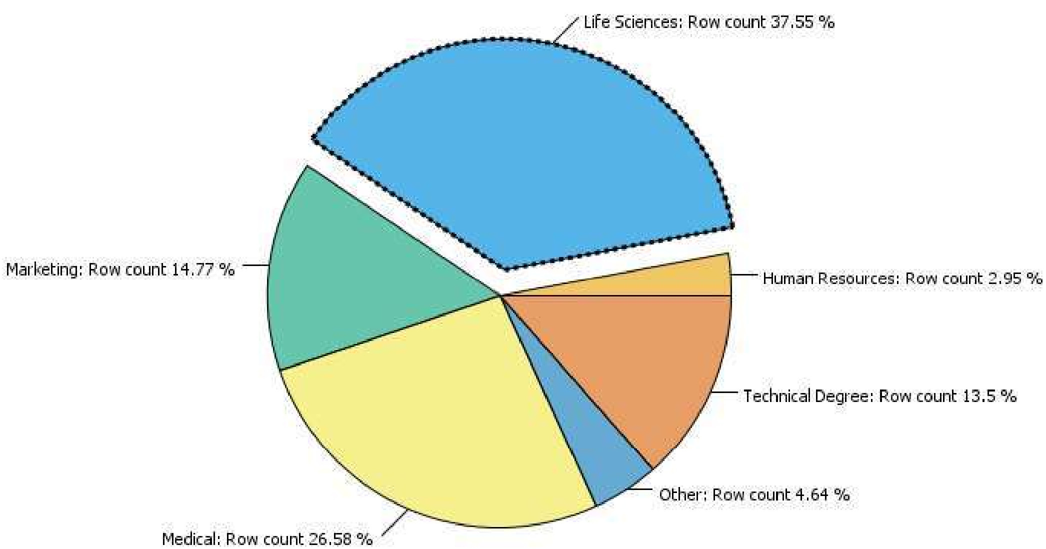


Fig. 5. Distribution of former employees by education field.

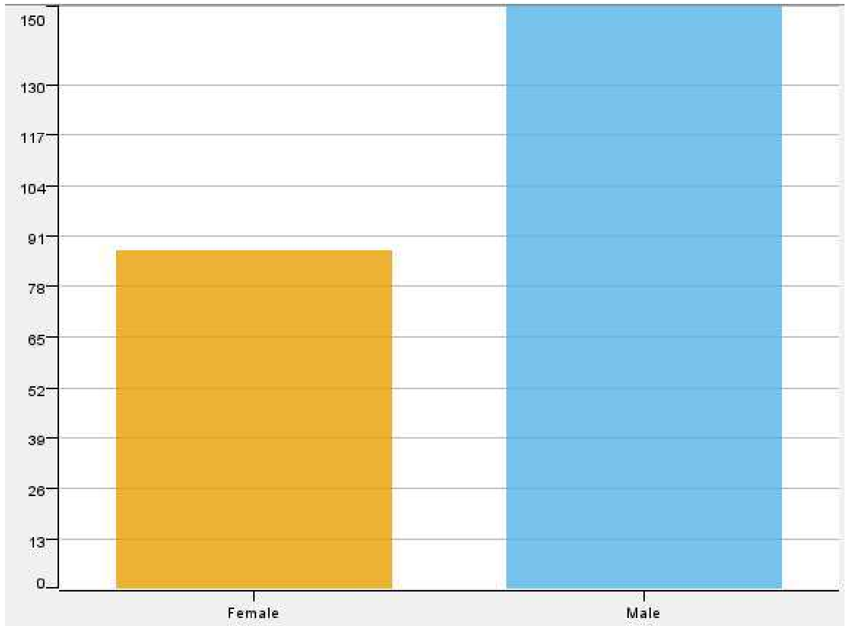


Fig. 6. Distribution of former employees by gender.

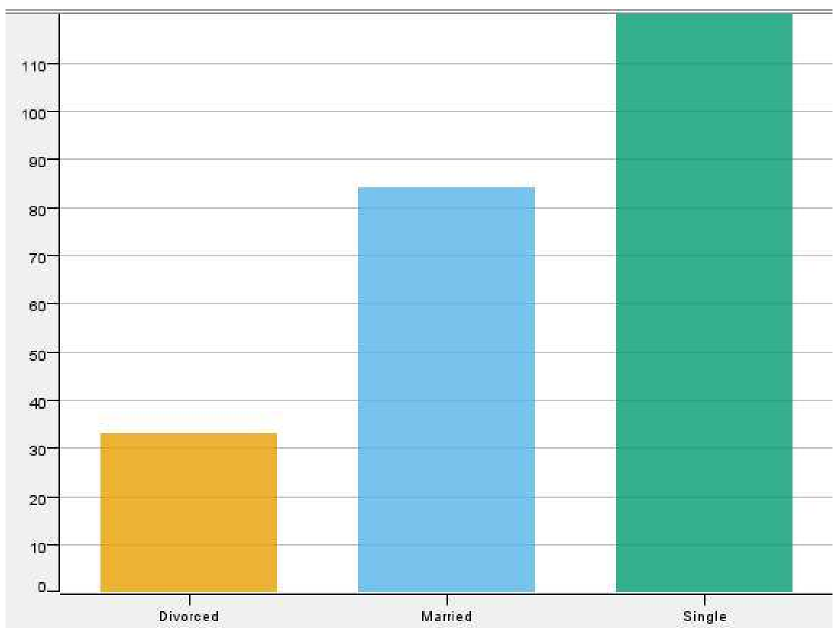


Fig. 7. Distribution of former employees by marital status.

- Employees' satisfaction rates: a ranking is associated to all employee's satisfaction/involvement and work-life balance variables, such as 1 'High', 2 'Medium', 3 'Low' and 4 'Very Low'. 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 (showing in Fig. 8. Distribution of former employees by work-life balance score).

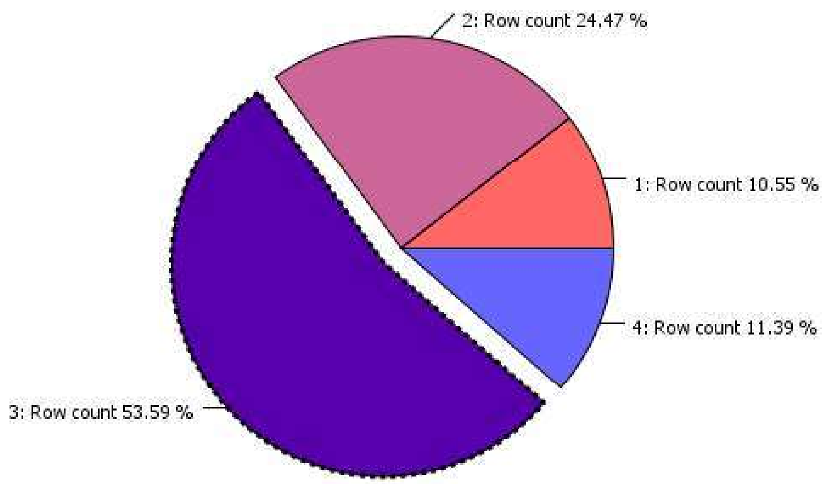


Fig. 8. Distribution of former employees by work-life balance score.

- 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 business travels numbers or duration of business travel, but data shows, that there is 88% of the leavers, who were having business travels (showing in Fig. 9. Distribution of former employees by business travels frequency). 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. This is showing Scatter Matric in Fig. 10. Distribution of former employees by monthly income and percent salary hike.

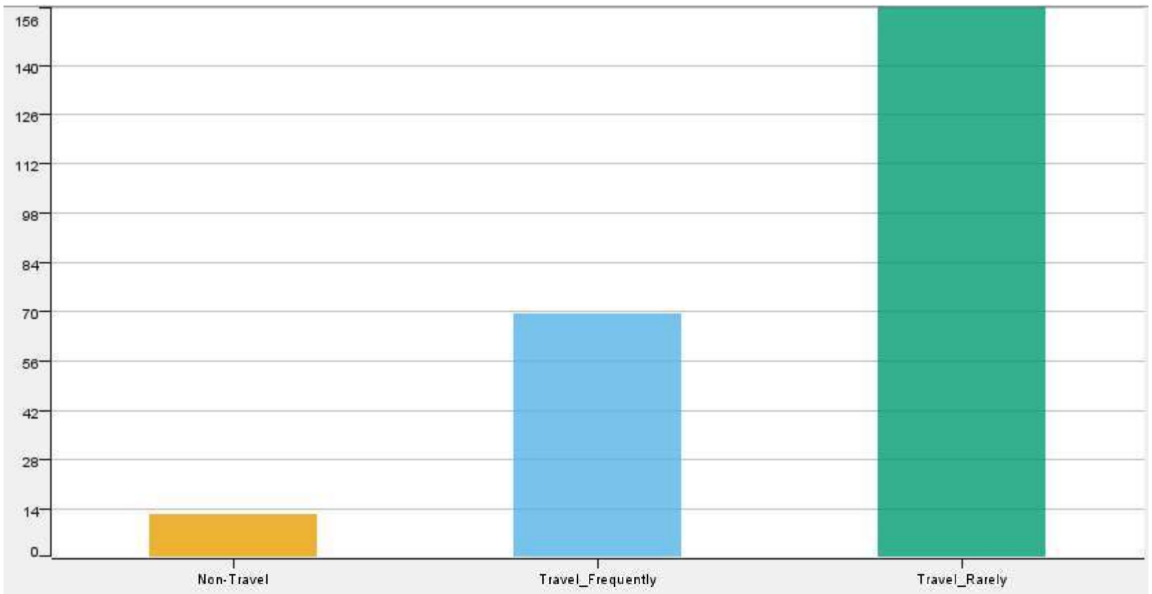


Fig. 9. Distribution of former employees by business travels frequency.

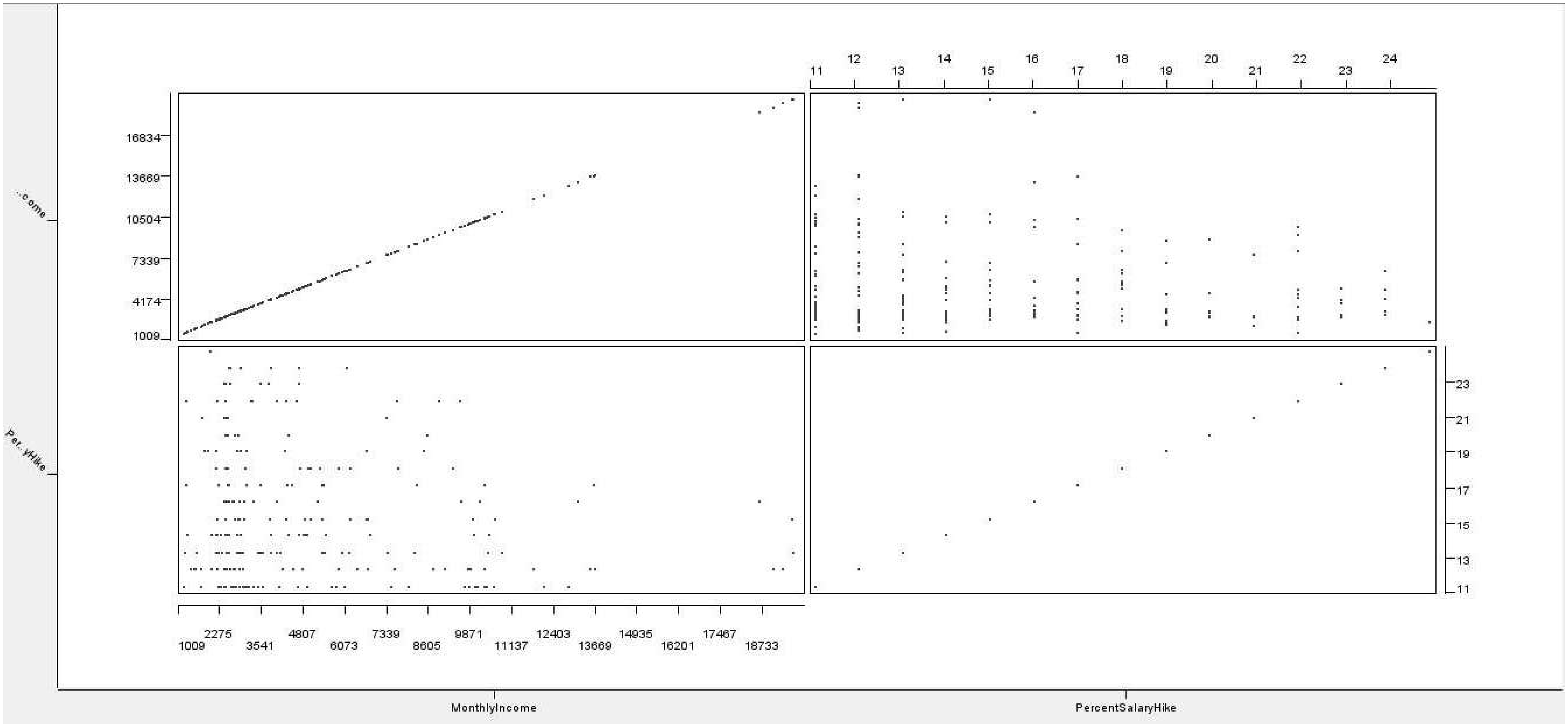


Fig. 10. Distribution of former employees by monthly income and percent salary hike.

- 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 (60,34%) (showing in Fig. 11. Distribution of former employees by employee level). Moreover, there is 43% of the employees, who no longer works for the company after working there for less than three years (showing in Fig. 12. Distribution of former employees by duration of work).

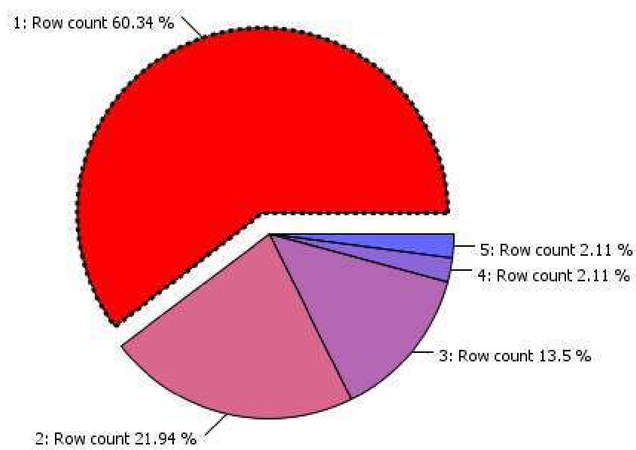


Fig. 11. Distribution of former employees by employee level.

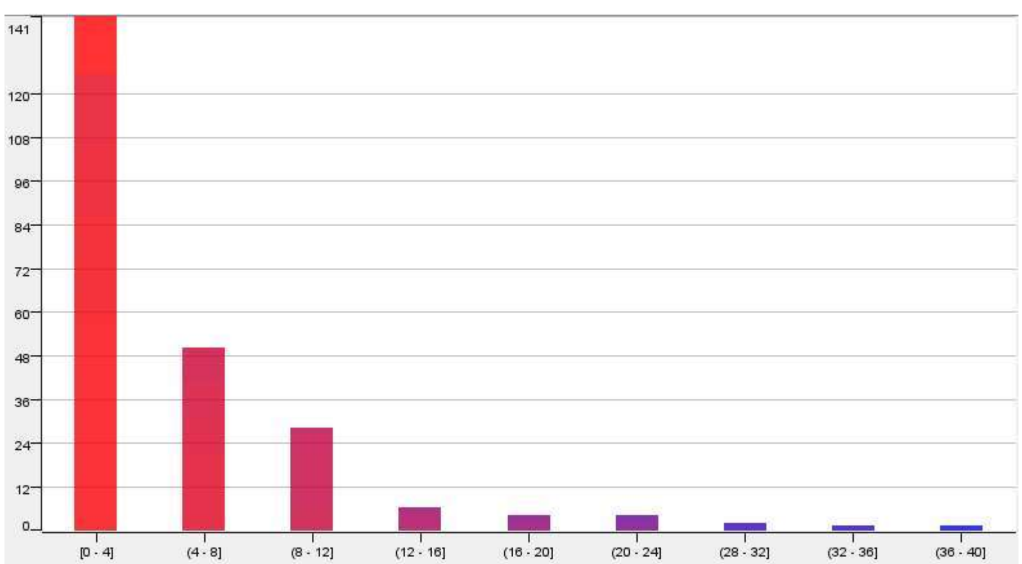


Fig. 12. Distribution of former employees by duration of work.

To extend the statistical analysis and to get more comprehensive overview, 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 Fig. 13. Employee attrition correlation matrix.

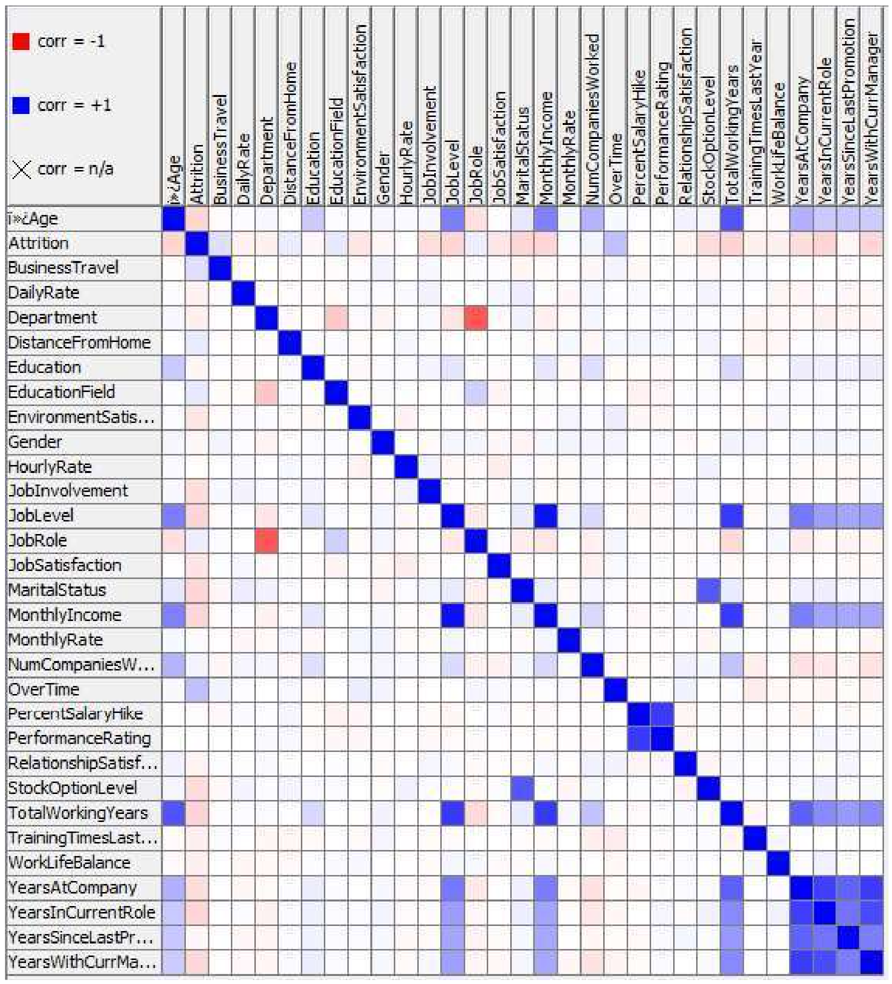


Fig. 13. Employee attrition correlation matrix.

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). 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 (Fig. 14. Variables correlation to employee attrition) defines *p-value* 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...	0.0955252620565 1235
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...

Fig. 14. Variables correlation to 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.

2.2.3. Machine learning algorithms.

In the current research, 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. Thereafter, Decision Tree, SVM and Naïve Bayes learner nodes have been applied to induce a classification decision in model's main memory. Along with learner nodes, predictors were used to get employee attrition predictions as the output. For this reason, in those three predictors, the nominal attribute "Attrition" served as the target attribute.

Decision Tree Learner 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 algorithm stands out from others since it can clearly observe the tendency of the dependent variable in tree structure and empowers the overview of the data in a tree structure (showing in Fig. 15. Decision Tree view for overtime). Tree structure view is available for both – learner and predictor nodes. With this model, 14% of employees were identified as potential leavers.

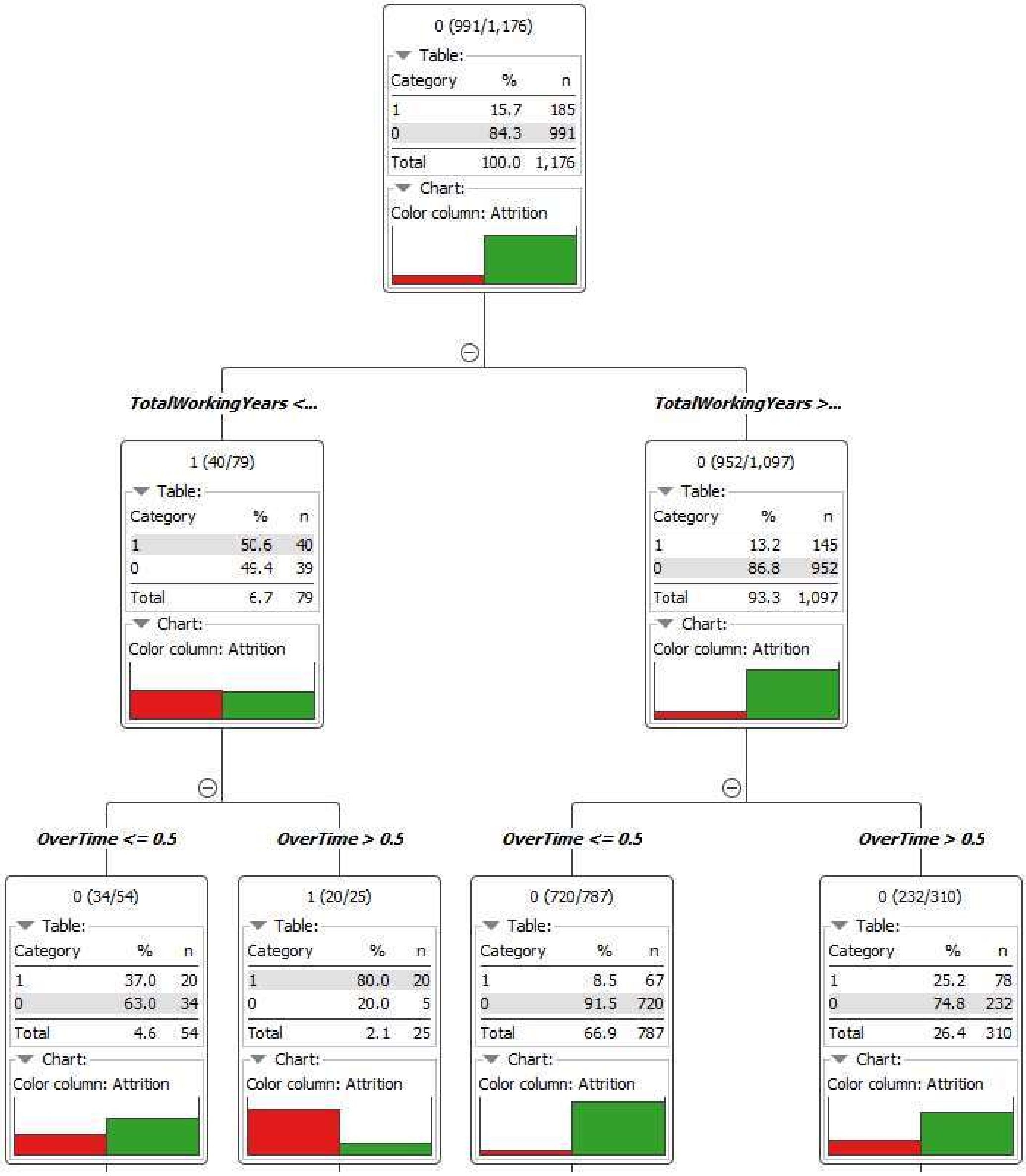
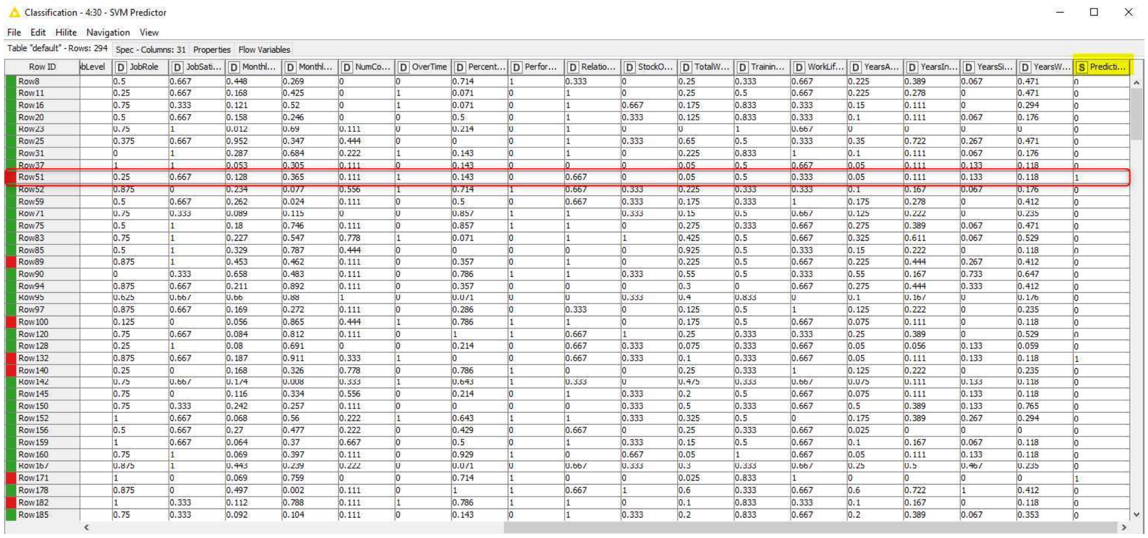


Fig. 15. Decision Tree view for overtime.

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 (see in Fig. 16. SVM Predictor output table). With this model, 7% of employees were identified as potential leavers.



Row ID	bLevel	JobRole	JobSati...	Monthl...	Monthl...	NumCo...	OverTime	Percent...	Perfor...	Relatio...	StockO...	TotalW...	Trainin...	WorkLif...	YearsA...	YearsIn...	YearsSi...	YearsW...	Predicti...
Row8		0.5	0.667	0.448	0.269	0	0	0.714	1	0.333	0	0.25	0.333	0.667	0.225	0.389	0.067	0.471	0
Row11		0.25	0.667	0.168	0.425	0	1	0.071	0	1	0	0.25	0.5	0.667	0.225	0.278	0	0.471	0
Row16		0.75	0.333	0.121	0.52	0	1	0.071	0	1	0.667	0.175	0.833	0.333	0.15	0.111	0	0.294	0
Row20		0.5	0.667	0.158	0.246	0	0	0.5	0	1	0.333	0.125	0.833	0.333	0.1	0.111	0.067	0.176	0
Row23		0.75	1	0.012	0.69	0.111	0	0.214	0	1	0	0	1	0.667	0	0	0	0	0
Row25		0.375	0.667	0.952	0.347	0.444	0	0	0	1	0.333	0.65	0.5	0.333	0.35	0.722	0.267	0.471	0
Row31		0	1	0.287	0.684	0.222	1	0.143	0	1	0	0.225	0.833	1	0.1	0.111	0.067	0.176	0
Row37		1	1	0.053	0.305	0.111	0	0.143	0	0	0	0.05	0.5	0.667	0.05	0.111	0.133	0.118	0
Row51		0.25	0.667	0.128	0.365	0.111	1	0.143	0	0.667	0	0.05	0.5	0.333	0.05	0.111	0.133	0.118	1
Row52		0.875	0	0.234	0.077	0.556	1	0.714	1	0.667	0.333	0.225	0.333	0.333	0.1	0.167	0.067	0.176	0
Row59		0.5	0.667	0.262	0.024	0.111	0	0.5	0	0.667	0.333	0.175	0.333	1	0.175	0.278	0	0.412	0
Row71		0.75	0.333	0.089	0.115	0	0	0.857	1	1	0.333	0.15	0.5	0.667	0.125	0.222	0	0.235	0
Row75		0.5	1	0.18	0.746	0.111	0	0.857	1	1	0	0.275	0.333	0.667	0.275	0.389	0.067	0.471	0
Row83		0.75	1	0.227	0.547	0.778	1	0.071	0	1	1	0.425	0.5	0.667	0.325	0.611	0.067	0.529	0
Row85		0.5	1	0.329	0.787	0.444	0	0	0	0	0	0.925	0.5	0.333	0.15	0.222	0	0.118	0
Row89		0.875	1	0.453	0.462	0.111	0	0.357	0	1	0	0.225	0.5	0.667	0.225	0.444	0.267	0.412	0
Row90		0	0.333	0.658	0.483	0.111	0	0.786	1	1	0.333	0.55	0.5	0.333	0.55	0.167	0.733	0.647	0
Row94		0.875	0.667	0.211	0.892	0.111	0	0.357	0	0	0	0.3	0	0.667	0.275	0.444	0.333	0.412	0
Row95		0.625	0.667	0.66	0.88	1	0	0.071	0	0	0.333	0.4	0.833	0	0.1	0.167	0	0.176	0
Row97		0.875	0.667	0.169	0.272	0.111	0	0.286	0	0.333	0	0.125	0.5	1	0.125	0.222	0	0.235	0
Row100		0.125	0	0.056	0.865	0.444	1	0.786	1	1	0	0.175	0.5	0.667	0.075	0.111	0	0.118	0
Row120		0.75	0.667	0.084	0.812	0.111	0	1	1	0.667	1	0.25	0.333	0.333	0.25	0.389	0	0.529	0
Row128		0.25	1	0.08	0.691	0	0	0.214	0	0.667	0.333	0.075	0.333	0.667	0.05	0.056	0.133	0.059	0
Row132		0.875	0.667	0.187	0.911	0.333	1	0	0	0.667	0.333	0.1	0.333	0.667	0.05	0.111	0.133	0.118	1
Row140		0.25	0	0.168	0.326	0.778	0	0.786	1	0	0	0.25	0.333	1	0.125	0.222	0	0.235	0
Row142		0.75	0.667	0.174	0.008	0.333	1	0.643	1	0.333	0	0.475	0.333	0.667	0.075	0.111	0.133	0.118	0
Row145		0.75	0	0.116	0.334	0.556	0	0.214	0	1	0.333	0.2	0.5	0.667	0.075	0.111	0.133	0.118	0
Row150		0.75	0.333	0.242	0.257	0.111	0	0	0	0	0.333	0.5	0.333	0.667	0.5	0.389	0.133	0.765	0
Row152		1	0.667	0.068	0.56	0.222	1	0.643	1	1	0.333	0.325	0.5	0	0.175	0.389	0.267	0.294	0
Row156		0.5	0.667	0.27	0.477	0.222	0	0.429	0	0.667	0	0.25	0.333	0.667	0.025	0	0	0	0
Row159		1	0.667	0.064	0.37	0.667	0	0.5	0	1	0.333	0.15	0.5	0.667	0.1	0.167	0.067	0.118	0
Row160		0.75	1	0.069	0.397	0.111	0	0.929	1	0	0.667	0.05	1	0.667	0.05	0.111	0.133	0.118	0
Row167		0.875	1	0.443	0.239	0.222	0	0.071	0	0.667	0.333	0.3	0.333	0.667	0.25	0.5	0.467	0.235	0
Row171		1	0	0.069	0.759	0	0	0.714	1	0	0	0.025	0.833	1	0	0	0	0	1
Row178		0.875	0	0.497	0.002	0.111	0	1	1	0.667	1	0.6	0.333	0.667	0.6	0.722	1	0.412	0
Row182		1	0.333	0.112	0.788	0.111	1	0.786	1	1	0	0.1	0.833	0.333	0.1	0.167	0	0.118	0
Row185		0.75	0.333	0.092	0.104	0.111	0	0.143	0	1	0.333	0.2	0.833	0.667	0.2	0.389	0.067	0.353	0

Fig. 16. SVM Predictor output table.

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

$$P\ (y\ /\ x) = \frac{P\left(\frac{x}{y}\right) * P(y)}{P\ (x)} \qquad (2)$$

where:

- x is the attribute.
- y is the class.
- P (y /x) 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. The output table adds an additional column for prediction and is presented in the same way as the SVM Predictor (see in Fig. 16. SVM Predictor output table). However, Naïve Bayes learner view provides concluded information for employee attrition class value distribution in additional table. The part of the table is showing in Fig. 17. Gaussian distribution for BusinessTavel variable per class. With this model, 31% of employees were identified as potential leavers.

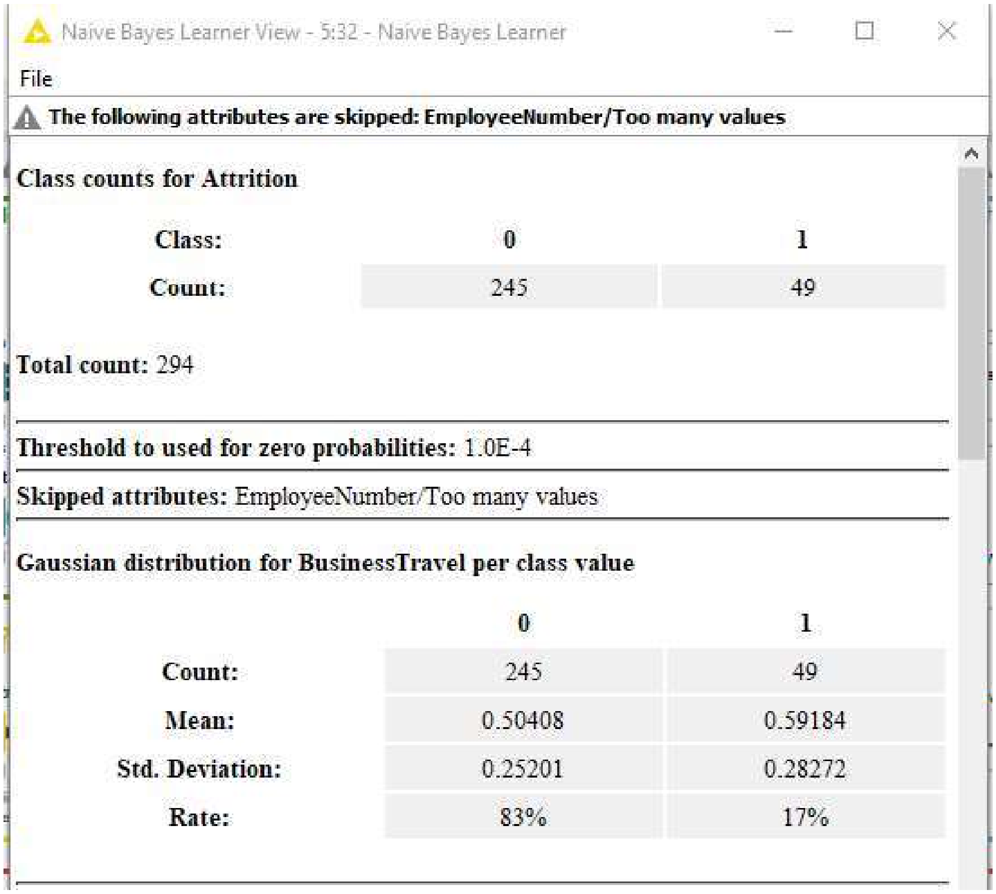


Fig. 17. Gaussian distribution for BusinessTavel variable per class.

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 4. 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).

Table 4. 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		

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, because this algorithm has been identified as the most reliable. Model consists of 5 stages: a) data collection, b) data exploration, c) data processing and feature engineering, d) data visualisation and e) Machine Learning algorithms. Model overview is showing in Fig. 18. Employee attrition Machine Learning (ML) predictive model).

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.

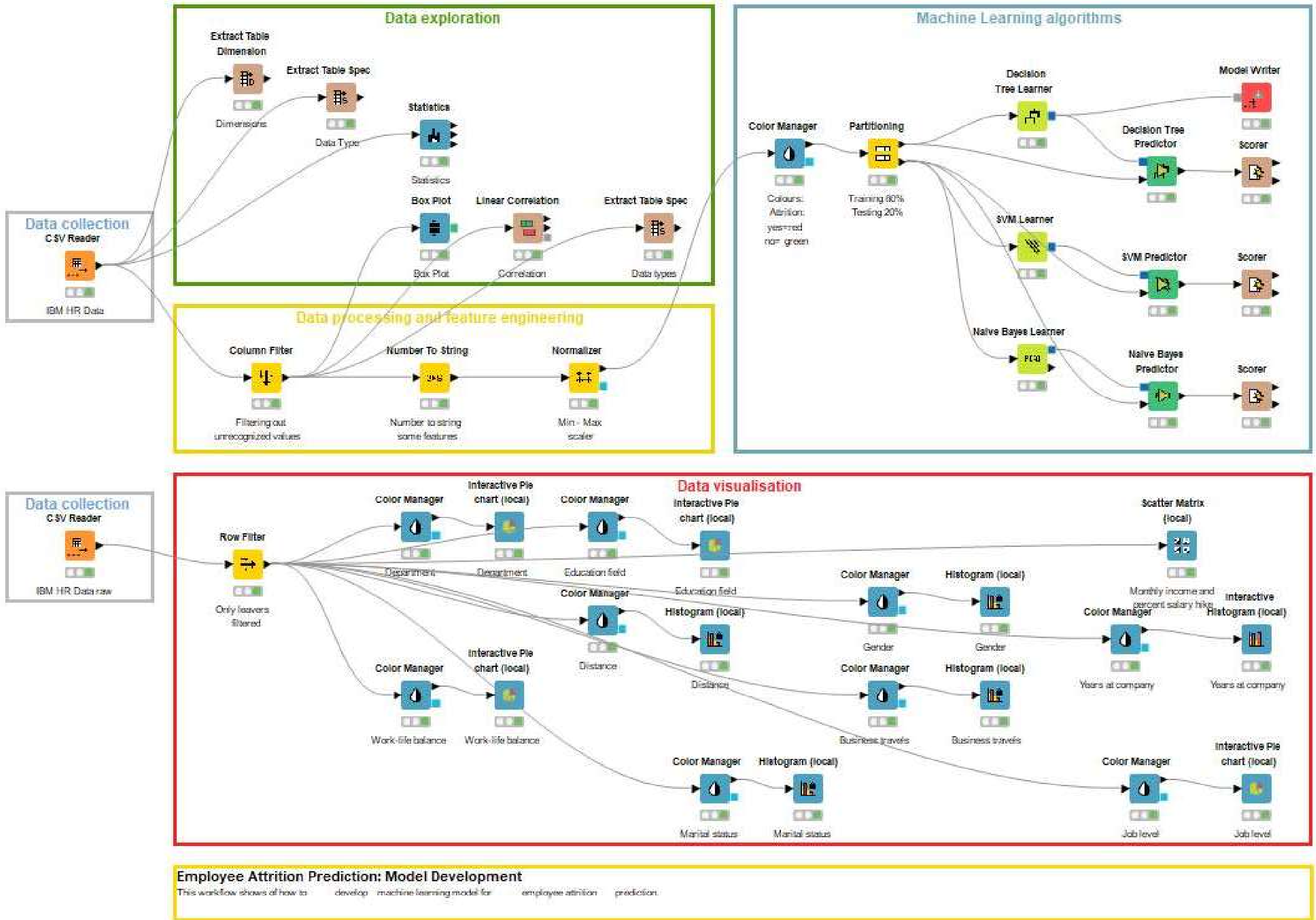


Fig. 18. Employee attrition Machine Learning (ML) predictive model

3. STRATEGIC DEPLOYMENT OF EMPLOYEE ATTRITION PREDICTION MODEL

3.1. Strategic employee attrition management

Based on narrative and systematic literature review, none of the previous employee attrition prediction models were described along with a strategical implementation plan. Therefore, with this research, the aim is to propose employee attrition predictive model together with strategic implementation in international company. Since company strategy is linked to process performance management, where company operations may be measured and monitored, it is critical to establish appropriate measurements. And thereafter, this research will explain how to build proper measurement for employee attrition and manage it strategically.

Based on Phillips J. J. & O. Connell A., (2004), job satisfaction could be a proper measure to identify employee attrition, because low job satisfaction rate may be translated into employee attrition rate. However, job satisfaction data alone could not be a leading indicator, so the key measure should be employee's intention to leave their position. Or in other words, employee attrition must be set as the key performance indicator. Authors have broadly explained how to manage employee attrition rate strategically, and broad concept of strategic employee attrition accountability approach, outlined in Fig. 19. Strategic accountability approach.

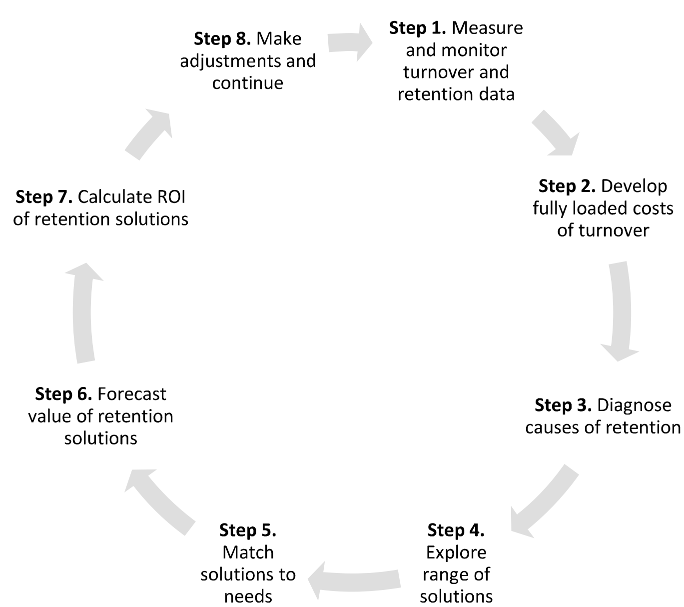


Fig. 19. Strategic accountability approach.

Source: Phillips J. J. & O. Connell A., (2004) With these steps, authors describe a concept of strategic accountability approach, which could be applied in the organisations working strategically with employee attrition. This concept is provided not as just one time deliverable, but more like as an ongoing process, which should take place as much as the company itself wants to improve its own employee attrition rate (Phillips J. J. & O. Connell A., 2004). Moreover, it is crucial to have continuously processes to ensure employee attrition targets. Organisations should seek to fulfilling employee's expectations and thus, create an impact on the performance of employee, which directly affects the organisation's performance. Organisation and employees are both dependent on each other to fulfil their goals and objectives (Bedarkar M. & Pandita D, 2014).

The steps outlined below is recommendation for international companies on how to manage employee attrition strategically. These steps are based on Phillips J. J. & O. Connell A., (2004) strategic accountability approach concept. The current study will rely on these steps to deploy strategic management for employee attrition by predicting it.

Step 1. Measure and monitor employee attrition data. This step explains that the value of the data should include the total rate of employee attrition, as well as the reasons for it. It is critical to set the targets or triggers for subsequent actions.

Step 2. According to authors, the following step should be to develop fully loaded costs of employee attrition. This step describes the financial impact of employee turnover. Capturing all costs associated with employee attrition may be difficult, but it is possible to calculate direct costs. Employee exit costs (administrative time, management time, benefit termination) may differ from replacement (planning recruitment, employment) and onboarding costs (training).

Step 3. Diagnose causes of employee attrition. With this step, authors recommend analysing causes of employees' attrition using questionnaires, surveys, or interviews. 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.

Step 4. Explore range of solutions. This step delves into the critical issue of employee salary and recognition. Standard salaries and increases, bonuses, benefits, rewards, and recognition must all be measured and decided.

Step 5. Match solutions to needs. This step is required to focus on solutions with the highest payoffs rather than attempting to tackle a wide range of issues.

Step 6. Forecast value of employee attrition solutions. Return on investment (ROI) calculations are possible at several different stages even before the employee attrition solution is implemented. However, to forecast the impact of this solution is an important issue when the reasons for the need for a forecasted ROI are examined.

Step 7. Calculate ROI of employee attrition solutions. To be useful, ROI processes must balance several factors, including feasibility, simplicity, credibility, and soundness.

Step 8. Adjust and continue. After the adjustment's implementation, communication and monitoring ensures information flows and that all stakeholders are aware of the successful solution.

Based on strategic accountability approach concept analysis, it was first mentioned, that it is essential to maintain and analyse employee attrition data since it provides useful insights into the reasons of attrition. With the insights, it would be possible to put them to use in trying to keep potential leavers. The target focus, of course must be the most valuable employees instead of trying to retain all the potential leavers. For this, employee attrition costs may be determined by calculating all the costs, spent to replace leaving employee. This is how the organisations can identify the highest employee attrition costs as well as the most valued employees. Managers or HR representatives should implement a solution and to apply it to identified potential company leavers and thus to improve working conditions. However, forecasting and calculating ROI on this solution is critical to strike a balance between feasibility, and credibility.

3.2. Predictive model deployment for strategic employee attrition management

As a takeaway from this research, employee attrition predictive model deployment will be applied for strategic employee attrition management. The below is recommendation, which could be applied for any kind of international company for the purpose to reduce employee attrition.

First, a KPI for employee attrition may be defined in the organisation to measure employee attrition data. Implementing KPI for employee attrition is critical for measuring, managing, and enhancing employee retention. It gives for organisations valuable information and aids in the creation of a more welcoming work environment to retain talent. Setting up employee attrition KPIs is a continuous process that necessitates DDD making and a long-term commitment to improve employee attrition. To monitor employee turnover, the employee attrition predictive model, developed earlier in this study could be used. Moreover, by assisting companies in proactively identifying and understanding the factors that contribute to attrition, a predictive model can be a powerful tool. Employee attrition predictive model can help organisations to understand the underlying causes and applying focused tactics to reduce attrition, retain talent, and eventually improve overall organisational performance. 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. The model can be re-trained by uploading new data.

To deploy the model and generate new predictions, earlier developed model in KNIME was re-used with Model Reader node (see in Fig. 20. Employee attrition prediction model deployment). This node receives an Employee attrition Machine Learning (ML) predictive model (see in the Fig. 18. Employee attrition Machine Learning (ML) predictive model), from a file created by the Model Writer node.

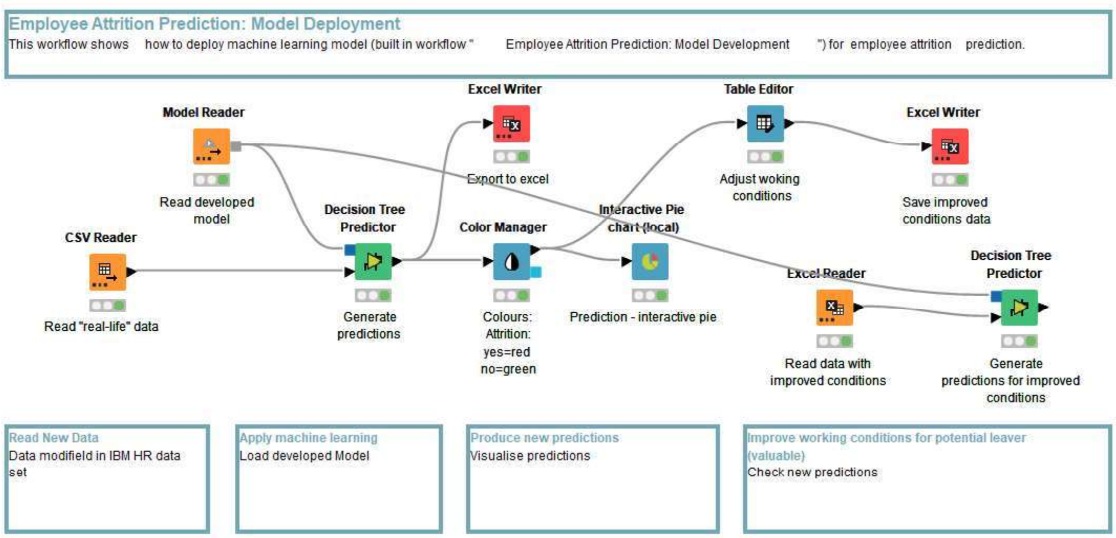


Fig. 20. Employee attrition prediction model deployment.

In parallel, there was CSV Reader node used for model deployment which reads new employee's data for predictions. To replicate real-life scenario and deploy the model, the same IBM HR data has been modified and uploaded with this node, leaving 10 employees' data to receive new employee attrition predictions. The Decision Tree algorithm, which was selected as the most reliable (with 97,4% accuracy), was used with the Decision Tree Predictor node to generate new predictions. These predictions are identification of the employee which is more likely to leave the organisation. When predictions for new employees' have been applied, colours were assigned to numeric values (when attrition value is "yes", then it is indicated as red, and when value is "no", then it is indicated as green). In other words, employees predicted as potential leavers are indicated with red colour. New predictions have been visualised in pie chart, using Interactive Pie chart (local) node (see in Fig. 21 Predictions for new applied employees' data). The model predicts, that 6 from 10 employees might be potential leavers.

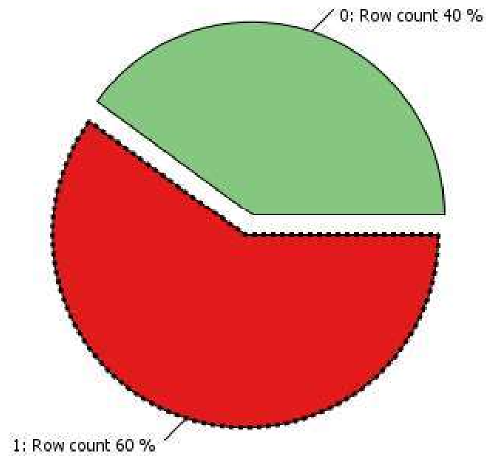


Fig. 21. Predictions for new applied employees' data.

It is possible to identify possible leavers by employee number, which could be found using the same node, extracting the table view (see in Fig. 22. Identification of potential leavers):

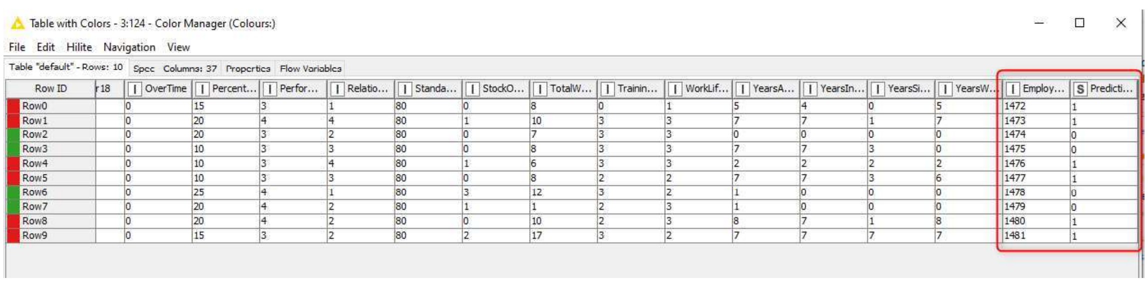


Table with Colors - 3:124 - Color Manager (Colours:)

File Edit Hilite Navigation View

Table "default" - Rows: 10 Spec Columns: 37 Properties Flow Variables

Row ID	r18	OverTime	Percent...	Perfor...	Relatio...	Standa...	StockO...	TotalW...	Trainin...	WorkLif...	YearsA...	YearsIn...	YearsSi...	YearsW...	Employ...	Predicti...
Row0		0	15	3	1	80	0	8	0	1	5	4	0	5	1472	1
Row1		0	20	4	4	80	1	10	3	3	7	7	1	7	1473	1
Row2		0	20	3	2	80	0	7	3	3	0	0	0	0	1474	0
Row3		0	10	3	3	80	0	8	3	3	7	7	3	0	1475	0
Row4		0	10	3	4	80	1	6	3	3	2	2	2	2	1476	1
Row5		0	10	3	3	80	0	8	2	2	7	7	3	6	1477	1
Row6		0	25	4	1	80	3	12	3	2	1	0	0	0	1478	0
Row7		0	20	4	2	80	1	1	2	3	1	0	0	0	1479	0
Row8		0	20	4	2	80	0	10	2	3	8	7	1	8	1480	1
Row9		0	15	3	2	80	2	17	3	2	7	7	7	7	1481	1

Fig. 22. Identification of potential leavers.

Afterwards, the data has been exported to Excel file using Excel Writer node. This node saves the input data table as an Excel spreadsheet, which may then be viewed with Microsoft Excel. Exported data might be used either for reporting, or further analysis.

As a result, from the above-mentioned example, there was identified 6 potential leavers from 10 employees in the company. However, it may not always be possible to enhance working conditions for all potential leavers. Therefore, the most detrimental replacement must be assessed by calculating the expenses associated with employee's replacement. Employee attrition costs might be computed as follows: (empty position coverage cost + cost to fill the empty position + onboarding and orientation expenses + productivity ramp-up cost) x number of employees lost in that position each year x 12 = yearly employee attrition expenses (Fortin D.), see equation below:

$$Yearly\ employee\ attrition\ expenses = (\text{coverage and position fill costs} + \text{onboarding expenses} + \text{productivity cost}) \times \text{no of lost employees} \times 12 \quad (3)$$

Different tools can be used to calculate these costs. Moreover, integrating with software applications that may offer professional solutions (for data storage, visualization, and so on) would be advantageous. However, in this study, an example Excel calculator was built to serve as a reference for the calculator. The example of employee attrition expenses might be found in the below table (Fig. 23. Employee attrition cost estimation tool). This example is created by taking one of the employee's data. Here is a short explanation what data might be calculated.

A. Benchmark Employee Cost. Benchmark Employee Cost here refers to labour-related expenses, as employee salary, benefits, bonuses etc. In this case, only the employee wage is mentioned.
B. Vacant Position Coverage Cost. This cost refers to the expenses spent by an organisation when position goes unfilled or empty for an extended length of time. This cost is calculated per the following formula: =IF(B9="","",B7 (no of day to fill the vacant position) * B8 (cost of covering a vacant position per day)).
C. Cost To fill the Position. This cost refers to an organisation's overall expenses involved in the process of recruiting, hiring, and onboarding a new employee to fill a specific position. This cost might vary based on the position, sector, the organisation's recruitment, and onboarding methods. This cost is computed by totalling HR work costs: =SUM(C15 (cost screening) : C16 (cost of interviews)).
D. Cost of Training. This cost refers to the expenses incurred by an organisation while giving training and development to a new employee to assist them in acquiring the necessary skills and knowledge for their job role. This cost can vary based on the role's complexity, the training methods used, and the length of the training program. This cost is calculated per the following formula: =IF(D20="","", D18 (Cost of Trainer Per Day) * D19 (Training Days)).
E. Productivity Cost. When hiring a new employee, productivity cost refers to the productivity loss incurred by the organisation and its existing employees because of onboarding and integrating the new employee into the workforce. This cost is related to a temporary drop in overall team or department efficiency while the new person

learns their function and becomes fully productive. This cost is calculated per the following formula: =IF(E24="","", E22 (salary per day) *E23 (working days during probation period / 2).

F. As per earlier provided formula (2), summing up B9 (total cost to cover vacant position), C17 (total cost to fill a vacant position), D20 (total cost of training) and E24 (total cost of productivity), would give total turnover cost per employee value.

Description				Amount
A.	Benchmark Employee Cost			
1	Leaving Employee's Annual Basic Salary			62844
2	Monthly Salary			5237
3	Working Days Per Year			230
4	Working Hours Per Day			8
5	Salary Per Day			273.23
6	Salary Per Hour			34.15
B.	Vacant Position Coverage Cost			
7	No of Day To Fill The Vacant Position			14
8	Cost of Covering A Vacant Position Per Day			200
9			Total Cost To Cover Vacant Position	2800
C.	Cost To Fill the Post			
10	Annual Salary Of Hiring Staff			30000
11	Working Days Per Year			230
12	Working Hours Per Day			8
13	Salary Per Day			130.43
14	Salary Per Hour			16.30
15	Cost of Screening			200
	Required Hours	Required Days		
	20	2.5		
16	Cost Of Interviews			100
	Required Hours	Required Days		
	10	1.25		
17			Total Cost To Fill a Vacant Position	300
D.	Cost of Training			
18	Cost of Trainer Per Day			273
19	Training Days			10
20			Total Cost of Training	2730
E.	Productivity Cost			
21	Annual Salary Of New Employee			62844
22	Salary Per Day			273
23	Working Days During Probation Period			58
24			Total Cost of Lost Productivity	7924
F.			Turnover Calculations	
25			Total Turnover Cost Per Employee	**13754**

Fig. 23. Employee attrition cost estimation tool.

According to calculated costs, organisations may assess the highest employee attrition costs as well as the most valuable employee. The costs could be compared as per the table in the example below (see Fig. 24. Employee attrition cost comparison).

Employee number	Prediction (Attrition)	Employee Attrition costs
1472	**1**	**15288**
1473	**1**	**13532**
1474	0	7352
1475	0	7020
1476	**1**	**10147**
1477	**1**	**9342**
1478	0	8530
1479	0	8575
1480	**1**	**22483**
1481	**1**	**13754**

Fig. 24. Employee attrition cost comparison.

As a result, employee “1480” has the greatest employee attrition costs, and as a result, preventive actions should be applied to enhance working conditions for this high-risk employee. These preventive actions could be considered as actions in the strategic employee attrition plan.

The KNIME tool was used to possibly enhance working conditions for the possible leaver. Data was manipulated with Table Editor node. This node displays data and provides possibility to edit the data selecting interactive view with JavaScript Table Editor. The following working conditions have been enhanced for potential leaver, who has been recognized as having the highest employee attrition costs: years since last promotion was changed to “0”, percent salary hike was increased from 20% to 25% and monthly income was increased to 11 907, adding up additional 25% to the current salary.

Data was exported to Excel and input with the Excel Reader node again to obtain a new prediction for potential leaver with number “1480” after changing working conditions. The Decision Tree algorithm was used to generate new forecast. As the result, the probability value of leaving the company has been changed to "0," which indicates that these efforts taken to enhance working conditions would result in the employee remaining in the organisation. At this stage, the involvement of management and HR representatives is critical to take the actions for potential leavers, or at the very least to have an employee performance talk.

Finally, employee attrition KPI should be monitored and measured constantly. Monitoring employee attrition KPI and making appropriate modifications is critical for an organisation’s effective performance management, decision-making, continuous improvement, accountability, and overall health and success. Since KPIs provide organisations with a defined framework for setting and tracking goals and driving their initiatives forward, they should stay as much as possible relative.

As the takeaway of this research, employee attrition predictive model was deployed for strategic employee attrition management. Moreover, a prototype of the model that allows proposing different proactive actions to retain employees and calculates the likelihood of attrition after those actions have been applied has been successfully tested. This was done by modifying the data and thus to improving working conditions for potential leaver. The results have been tested by re-calculating employee attrition probability. The results showed zero probability of leaving the company, because working conditions have been improved.

As a recommendation the following actions to reduce employee attrition rate ensuring continuity might be provided to business management: 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 KNIME ML employee attrition predictive model must be applied, as it is accurate tool to predict employee attrition. Because this model learns from the data and so delivers more accurate predictions, adding more data to the predictive model may enhance it. It is important to get as much as possible information about employee, job, and relationship satisfaction measurements. The aim of this tool is not only to predict potential leavers, but also to measure and monitor employee attrition KPI. It is important, that KPI should stay as much as possible relative. When predictions are generated, it may be analysed with the focus of the highest costs for position replacement. For this, employee turnover costs may be determined by calculating the following costs as per this research example: benchmark, position coverage, training, productivity cost, and costs to fill the position. And this is how the organisations can identify the highest employee attrition costs as well as the most valued employee. Managers or HR representatives should implement a strategic employee attrition plan, and to apply it to identified potential company leavers based on deployed model predictions. Preventive actions should be applied to improve working conditions for the most valuable potential leavers. This might be achieved by utilizing the same KNIME model and modifying the data until the probability becomes zero. A strategic employee attrition plan should assist in lowering the employee attrition rate and controlling the employee attrition KPI. This recommendation is not intended to be a one-time fix, but it may aid in reducing employee attrition and ensure, that the employee attrition target is consistently met.

RESEARCH LIMITATION

Since determining research limitations is critical for providing a balanced and accurate assessment of the study's findings, some limitations must be observed in this chapter. The main research limitations are related to resources and data. As the data set is the most essential component of machine learning, it is vital to highlight the most significant study limitation of restricted access to real-company data. Since machine learning creates the best accurate predictions with as much data as possible. This constraint may have an impact on the study's comprehensiveness and representation. However, obtaining such information would be difficult because most of the data variables are sensitive and confidential.

Moreover, no additional consultants were included in this research (except the research supervisor), and it could possibly affect the breadth of data analysis, methodological approaches, and the ability to address complexities within the study. AI and ML are currently widely researched concepts, and so there may be an endless investigation to identify the best solution to a certain business challenge. Technically, AI and ML concepts encompass a wide range of software and applications. As a result, broader research is likely to yield more and more superior solutions.

CONCLUSIONS AND RECOMMENDATIONS

1. PPM was discovered to be crucial for any business organisation during the narrative literature because it aids in the measurement, monitoring, and optimization of processes, hence assists in the elimination of bottlenecks, reduction of waste, and enhancement of productivity. It has been revealed that the relationship between PPM and employees is mutual because employee engagement, involvement, skills, and motivation have an impact on product and service quality. Orientation to employees and their loyalty is important, and ignoring employee-related issues may result in poor quality and productivity. Therefore, it is critical to strike a balance that aligns employee interests with PPM objectives. However, employee attrition has been disclosed to be increasing in organisations, thus resulting in both financial and business management losses.
2. AI is increasingly being utilized to analyse customers, develop products and services, and automate corporate operations, therefore it may be ideal for solving any business problems. The most common data science approaches nowadays are machine learning and data mining, but it is preferable to use a combination of AI techniques. There have been several earlier studies about employee attrition prediction using AI, however none of the previous models were produced in conjunction with strategical implementation. Most of the studies only reviewed ML techniques to solve this problem.
3. Based on the best practices during the earlier research, current predictive model has been developed with different stages: data collection, data exploration, data processing and feature engineering, data visualisation and Machine Learning algorithms. For this research an open dataset of IBM HR Analytics Employee Attrition & Performance, extracted from Kaggle, of 1470 employees was used. The applied data includes employees' basic information, demographics, employee satisfaction rates, working conditions, and personal experience. Age, education, business trip frequency, work-life balance, job involvement, environmental satisfaction, salary, job position, and years in company were discovered to be factors, that may influence employee attrition. The DTC, SVM, and Naïve Bayes algorithms were tested, and the DTC proved to be the most reliable, with 97,4% accuracy in predicting employees, who could leave the organisation. Since this is a self-learning algorithm, accuracy is expected to fluctuate as more datasets are added. DTC algorithm has been chosen for model deployment.
4. Additional literature analysis was conducted to support the strategic model deployment. Based on this analysis, a comprehensive strategic implementation plan has been defined together with predictive model prototype. Furthermore, a prototype of deployed model,

developed to suggest proactive actions for employee retention has undergone successful testing. After the implementation of these actions, the model computed a revised probability of employee attrition. The findings revealed a negative probability of leaving the company because working conditions have been improved. As per this research's recommendation, employee attrition should be established as organisational KPI first. Since KPIs must stay relative, they must be measured, monitored, and updated when needed. To generate forecasts of potential company leavers, the ML employee attrition predictive model should be applied. Employee attrition costs should be calculated to understand the largest damage of position replacement for potential leavers, as well as to identify the most valued employee. A strategic employee attrition plan must be developed and executed for identified potential company leavers through preventative actions to improve working conditions. This solution may aid in reducing employee attrition and ensuring that the employee attrition target is met on a consistent basis. In the long run, this approach would improve the others organisation's KPIs such as productivity, efficiency, and customer satisfaction.

REFERENCES

27 US Employee Turnover Statistics [2022]: Average Employee Turnover Rate, Industry Comparisons, And Trends – Zippia. Retrieved September 24, 2022, from https://www.zippia.com/advice/employee-turnover-statistics/

Al-Darraji, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., Abdulmalik, H. A. (2021). Employee Attrition Prediction Using Deep Neural Networks [Article]. *Computers (Basel)*, *10*(11), 141. https://doi.org/10.3390/computers10110141

Alsheref, F. K., Fattoh, I. E., M. Ead, W. (2022). Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms [Article]. *Computational Intelligence and Neuroscience*, *2022*, 1–9. https://doi.org/10.1155/2022/7728668

Alao, D., Adeyemo, A. B. (2013). ANALYZING EMPLOYEE ATTRITION USING DECISION TREE ALGORITHMS. *Information Systems & Development Informatics*, *4*(1).https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1012.2947&rep=rep1&type=pdf

Awwad, M. S., & Heyari, H. I. (2022). Predicting employee turnover using financial indicators in the pharmaceutical industry [Article]. *Industrial and Commercial Training*, *54*(3), 476–496. https://doi.org/10.1108/ICT-01-2022-0004

Balaban, N., Belić, K., Gudelj, M. (2011). Business Process Performance Management: Theoretical and Methodological Approach and Implementation. *Management Information Systems*, *6*(4). 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*.

Bedarkar M., Pandita D. (2014). A Study on the Dirvers of Employee Engagement Impacting Employee Performance. *Procedia – Social and Behavioral Sciences.* 133 (2014), 106-115. https://doi.org/10.1016/j.sbspro.2014.04.174

Berthold, M. R. (2020). *Guide to intelligent data science: How to intelligently make use of real data (2nd ed.)*. [Book]. Springer. (pp.2-23). https://doi.org/10.1007/978-3-030-45574-3

Bhardwaj, S., Singh, A., Lakshmi. (2016). FACTORS AFFECTING EMPLOYEE ATTRITION AMONG ENGINEERS AND NON ENGINEERS IN MANUFACTURING INDUSTRY. *Voice of Research*, *5*(2). ISSN 2277-7733. http://voiceofresearch.org/Doc/Sep-2016/Sep-2016_7.pdf

Bhatia, P. (2019). *Data mining and data warehousing: principles and practical techniques* [Book]. Cambridge University Press. (pp. 388-404).

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

Chyi Lee, C., Yang, J., & Ming Yu, L. (2001). The knowledge value of customers and employees in product quality [Article]. *The Journal of Management Development*, *20*(8), 691–706. https://doi.org/10.1108/02621710110401419

Clack, L. (2020). *Employee Engagement: Keys to Organisational Success*. *The Palgrave Handbook of Workplace Well-Being.* Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-02470-3_77-1

El-Rayes, N., Fang, M., Smith, M., Taylor, S. M. (2020). Predicting employee attrition using tree-based models [Article]. *International Journal of Organisational Analysis (2005)*, *28*(6), 1273–1291. https://doi.org/10.1108/IJOA-10-2019-1903

Fallucchi, F., Coladangelo, M., Giuliano, R., William, De Luca E. (2020). Predicting Employee Attrition Using Machine Learning Techniques [Article]. *Computers (Basel)*, *9*(4), 86. https://doi.org/10.3390/computers9040086

Fleischmann, A. (2012). *Subject-oriented business process management* [Book]. (pp. 207-225). Heidelberg: Springer. https://link.springer.com/content/pdf/10.1007/978-3-642-32392-8.pdf

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

Fortin D., How to calculate employee turnover cost. https://www.predictiveindex.com/blog/how-to-calculate-employee-turnover-cost/

Gartner, technological research, and consulting company for tool comparison. 2023, May data. https://www.gartner.com/reviews/market/multipersona-data-science-and-machine-learning-platforms/vendor/KNIME/product/KNIME-analytics-platform/alternatives

Gilula, M. (2016). *Structured search for big data: from keywords to key-objects* [Book]. Morgan Kaufmann. (pp. 1-48).

Gruman, J. A., & Saks, A. M. (2011). Performance management and employee engagement [Article]. *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

Harmon, P. (2014). *Business process change: a business process management guide for managers and process professionals* (3rd ed.) [Book]. Elsevier. *Business process change.* (pp. 1-21). https://www.sciencedirect.com/book/9780128003879/business-process-change

Hey, R. B. (2017). *Performance management for the oil, gas, and process industries: a systems approach. What is Performance Management?* (pp. 1-9). Cambridge. [Book]. Gulf Professional Publishing is an imprint of Elsevier. https://www.sciencedirect.com/book/9780128104460/performance-management-for-the-oil-gas-and-process-industries

Hutama, J. A. N., & Sagala, E. J. (2019). INFLUENCE OF EMPLOYEE ENGAGEMENT AND ORGANISATIONAL CULTURE TOWARDS EMPLOYEE PERFORMANCE [Article]. *Manajemen Bisnis (Online)*, *9*(2). https://doi.org/10.22219/jmb.v9i2.7939

Išoraitė, M. (2021). 7 P Marketing Mix Literature Review. International journal of trend in scientific research and development, 5(6), 1586-1591. https://www.ijtsrd.com/papers/ijtsrd47665.pdf

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

Kashif, M., Zarkada, A., & Thurasamy, R. (2017). Customer aggression and organisational turnover among service employees [Article]. *Personnel Review*, *46*(8), 1672–1688. https://doi.org/10.1108/PR-06-2016-0145

Kumar Jain, M. (2013). An Analysis of Marketing Mix: 7Ps or More. Asian Journal of Multidisciplinary Studies, 23-28. https://core.ac.uk/download/pdf/229673197.pdf

Lindsey Hall, K. K., Baker, T. L., Andrews, M. C., Hunt, T. G., & Rapp, A. A. (2016). The Importance of Product/Service Quality for Frontline Marketing Employee Outcomes: The Moderating Effect of Leader-Member Exchange (LMX) [Article]. *Journal of Marketing Theory and Practice*, *24*(1), 23–41. https://doi.org/10.1080/10696679.2016.1089762

Melcher, K., & Silipo, R. (2020). *Codeless deep learning with KNIME: build, train, and deploy various deep neural network architectures using KNIME analytics platform* (Community edition.) [Book]. Packt.

Muhsin, A., & Budaya, P. W. (2019). ANALYSIS OF EMPLOYEES' WORKLOAD QUALITY CONTROL [Article]. *Russian Journal of Agricultural and Socio-Economic Sciences*, *89*(5), 176–180. https://doi.org/10.18551/rjoas.2019-05.21

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

Peters, J. (2019). *Employee Engagement* [Book]. 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* [Book]. 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

Phillips J. J. & O. Connell A. (2004). *Managing Employee Retention.* [Book]. Taylor and Francis. https://doi.org/10.4324/9780080510873

Provost, F., & Fawcett, T. (2013). *Data science for business: [what you need to know about data mining and data-analytic thinking]* [Book]. O'Reilly. (pp.1-78).

Pousa, C., Liu, Y., & Aman, A. (2020). The effect of managerial coaching on salesperson's relationship behaviors: new evidence from frontline bank employees in China [Article]. *International Journal of Bank Marketing*, *38*(6), 1259–1277. https://doi.org/10.1108/IJBM-12-2019-0437

Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022a). Predicting Employee Attrition Using Machine Learning Approaches [Article]. *Applied Sciences*, *12*(13), 6424. https://doi.org/10.3390/app12136424

Ray, A. N., Sanyal, J. (2019). Machine learning based attrition prediction. *2019 Global Conference for Advancement in Technology (GCAT)*. (pp. 1-4). https://doi.org/10.1109/GCAT47503.2019.8978285

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

Simon, D. H., Gómez, M. I., McLaughlin, E. W., & Wittink, D. R. (2009). Employee attitudes, customer satisfaction, and sales performance: assessing the linkages in US grocery stores [Article]. *Managerial and Decision Economics*, *30*(1), 27–41. https://doi.org/10.1002/mde.1433

The people make the firm. (2022). *Strategic Direction (Bradford, England)*, *38*(12), 31–32. https://doi.org/10.1108/SD-10-2022-0124

Usha, P. M., & Balaji, N. V. (2021). A comparative study on machine learning algorithms for employee attrition prediction [Article]. *IOP Conference Series. Materials Science and Engineering*, *1085*(1), pp. 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 [Article]. *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* [Book]. Wiley-Blackwell. (pp. 1-9).

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques* (3rd ed.). (pp.3-145). [Book]. Elsevier/Morgan Kaufmann.

Yahia, N. ben, Hlel, J., & Colomo-Palacios, R. (2021). From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction [Article]. *IEEE Access*, *9*, 60447–60458. https://doi.org/10.1109/ACCESS.2021.3074559