UDC 004.13.8
DOI https://doi.org/10.32782/2663-5941/2025.3.2/55

Ushakova I.O.
Simon Kuznets Kharkiv National University of Economics

Bondarenko D.O.
Simon Kuznets Kharkiv National University of Economics

Chirva Yu.Ye.
Simon Kuznets Kharkiv National University of Economics

Znakhur L.V.
Simon Kuznets Kharkiv National University of Economics

COMPUTER MODELING OF COMPANY EMPLOYEE CHURN USING MACHINE LEARNING AND PREDICTIVE ANALYTICS METHODS

The study proposes an approach to management employee churn using machine learning, predictive analytics methods and with the support of information technologies. It systematically explores the theoretical underpinnings of employee turnover, categorizing it into various types and identifying key components for measurement, as well as factors influencing churn rates and their potential regulation. The discourse highlights the importance of employee retention strategies, emphasizing elements such as compensation, work planning, performance evaluations, training programs, and opportunities for career advancement. The authors propose utilizing decision tree and logistic regression methodologies to predict employee churn, selecting a binary classification criterion that distinguishes between employees who remain with the company and those who depart. Two predictive models are developed, showcasing significant accuracy metrics: the decision tree model demonstrates an impressive 91.3% accuracy on training data and 74.19% on test data, while the logistic regression model indicates 88.41% accuracy on training data and 90.32% on test data. These findings underscore the reliability of the proposed models in forecasting employee turnover. Furthermore, the article outlines practical applications for these models, providing actionable insights for organizations aiming to implement data-driven strategies for improving employee retention. By leveraging advanced analytical techniques, the study contributes valuable methodologies and frameworks for companies grappling with the challenges of employee churn, positioning itself as a critical resource for human resource professionals and organizational leaders seeking to enhance workforce stability and performance. Overall, the research underscores the potential of predictive analytics in informing strategic decisions regarding employee retention and maintaining a healthy climate in the company.

Key words: *computer modeling, machine learning, logistic regression, decision tree, employee churn, predictive analytics.*

Statement of the problem. One of the main tasks of modern companies is the search and selection of personnel, their retention in the company. HR departments of companies in the current conditions have to respond quickly to all changes in the specialist market and constantly process a large amount of information. In these conditions, conducting scientific research on computer simulation of employee churn is an extremely urgent task. Determining staff churn allows the management of an company to determine which part of the staff remains loyal to the company, and which part is able to go to competitors, thereby reducing profits. The company's profit depends on the quality and ease of forecasting the churn of personnel.

Timely forecasting of the churn of employee is the key to the successful and effective work of HR managers in the field of search, selection and retention of personnel, and therefore to the effective work of the company. In today's world, there are many mathematical methods and software tools that can be used to build predictive models that allow timely forecasting of staff churn and simplify business processes.

Analysis of recent research and publications. The problem of modeling the predictive analytics of the churn of company employee in economic and mathematical science was raised by a number of researchers. For example, scientific articles [1, 2] propose a method of substantiating the innovative strategy of a company and analyzing employee management mechanisms in the context of innovative business activities.

Researches [3–5] offer methods for optimizing employee management of companies to improve control processes and evaluate the effectiveness of web technologies. The issue of optimizing the interface of predictive analytics information systems is considered in scientific articles [6, 7].

Predictive analytics is also important in the analysis of employee churn. Predictive models based on classification and machine learning methods allow you to determine which class an employee belongs to: will be fired or will continue to work. This allows companies to provide decision support for creating sound and effective HR strategies [8–11]. Various methods are used to create predictive models, such as Logistic Regression, Decision Trees, Decision List Algorithm etc. [12–15].

In order to take competitive positions in the market, companies are increasingly using predictive analytics and information technology in HR management it in their activities. Innovative methods and technologies help companies identify factors and trends in employee churn in advance and develop appropriate measures to reduce employee churn. Therefore, there is a need to develop a methodology for predicting the churn of company employee using predictive analytical models, based on a reasonable set of factors affecting the outflow.

Task statement. The purpose of this article is to substantiate the technology of using mathematical methods and computer modeling tools for building and analyzing predictive models of employee churn.

When analyzing employee churn, it is important to assess which employee leave and which remain in the organization, whether the trend of changing the quality of employee coincides with the strategic settings defined by the management. Without this assessment, it is impossible to understand whether the existing level of variability is a positive or negative phenomenon. If it is precisely those employee who need to be got rid of, it means that the organization is on the right track. If it loses the best employee, then the issue of churn must be quickly resolved. Also, the situation when the changeability and rotation of personnel, both large and small, does not contribute to improving the quality of work is also not encouraging.

Outline of the main material of the study. It should be noted that the statement that the lower the churn rate, the better is false. It is not true. Extremely low churn rates indicate that the employee of the organization is not updated much and even ineffective employees do not leave it. The main issue of any company is the provision of uninterrupted work, avoiding problems related to the unexpected dismissal of employee. In modern theories of employee churn management, there are methods that allow you to calculate with a certain percentage of probability how much a person will work. They make it possible to determine this already in the process of negotiations and interviews, as well as based on the analysis of information about the dates of admission and dismissal in the candidate's employment book. It is useful to process them on a computer and get the result in graphic form. The degree of loyalty of the candidate to the previous employers and the tendency to change the attitude immediately becomes obvious. The HR manager can ask to comment on the graphs and observe the sincerity and logic of the answers, as well as the applicant's attitude, constructiveness, intellectual abilities and communication style. According to the form of manifestation, physical and hidden employee churn should be distinguished. Physical churn includes those employees who, due to various reasons, are fired and leave the organization. Hidden (psychological) staff churn occurs among those employees who outwardly do not leave the organization, but actually leave it, disengage from organizational activities.

The next division of employee churn is churn in relation to the organization – intra-organizational and inter-organizational churn. Internal organizational churn is related to labor movements within the organization, and external churn is between organizations, industries and spheres of the economy.

Depending on the size, liquidity is divided into natural and excess. Natural churn (3–5% per year) contributes to the timely renewal of the team and does not require special measures on the part of the management and employee service. High levels of staff churn lead to significant economic losses. Companies get organizational, personnel, technological and psychological difficulties as a result.

Almost all large organizations measure employee churn. At the same time, you should pay attention to the following components:

the general level of churn and its comparison with churn in other companies on the market (benchmarking);

dynamics of churn rate over time;

changing the structure of reasons for churn;

change in churn of different categories of personnel.

Companies participate in specially organized studies to obtain benchmarks or churn rates in the market or industry. In the process of research, employee-initiated churn and employer-initiated churn are often measured separately. And also the churn of new employee: during the first 3 months or during the first year of work.

To analyze employee churn, first of all, it is necessary to consider the factors affecting employee churn and their regulation. These include the following factors:

material (uncompetitive pay rates, unfair pay structure, unstable earnings);

organizational (schedule, regime, working conditions do not meet the employee's expectations, lack of opportunity for career growth, professional development);

interpersonal (relationships with management and colleagues);

the employee's age (the most risky age of transition to another job is under 25);

employee qualification (the lower the qualification, the more often they change jobs);

the employee's place of residence (the further the employee lives from work, the greater the risk of his leaving);

length of service at the enterprise (after three years of experience there is a sharp decrease in churn, which is explained both by the age factor and adaptation problems).

Retention measures should include each direction where adherence problems and dissatisfaction may appear such as:

payment;

work planning;

assessment of activity results;

training;

career development;

commitment and engagement;

team feeling;

conflicts.

To reduce staff churn, leading companies are now widely using predictive analytics, which is based on statistical methods, methods of intelligent data analysis, game theory, analysis of current and historical facts to make predictions about future events. Predictive analytics relies heavily on advanced information technologies for intelligent data analysis.

Logistic regression and decision trees are the most popular methods for predicting employees churn. The features of their possible application are analyzed below Logistic regression allows you to determine the predicted outcome for two conditions, using several factors. To determine employees churn, you need to divide them into two groups: employees either leave or stay in the company.

Decision trees are also an important tool in the work of every specialist involved in data analysis. This method has several algorithms for tree construction and is widely used to analyze staff churn. A decision tree allows you to build rules based on tracking with a certain probability all paths from the root node to the leaves of the tree. Both methods have certain advantages and disadvantages, therefore, the third section will implement models based on these methods with further comparative analysis of their obtained results.

Currently, the methods of predictive analytics are quite popular for solving the problem of employee churn. Predictive analytics is a data analysis method. The main task of these methods is to determine the state or direction of behavior of any entities in the future. The results obtained are used to justify management decisions [1–2, 6–7, 9, 14]. For forecasting, the following methods of data mining, statistical analysis, game theory, etc. can be used. Predictive analytics compares the current situation with historical data analytics, which makes it possible to determine predictions about future events. Thus, predictive analytics models in real business conditions determine the relationships between several indicators that influence decision-making, and allow assessing the risks associated with a certain set of factors. Today, predictive analytics relies heavily on advanced machine learning technologies. Data scientists use deep learning and complex algorithms to analyze multiple variables to create predictive models that can predict typical behavior based on big data. Next, we will consider methods that are widely used for predictive modeling, such as logistic regression and decision trees.

1. Logistic regression.

Let's define the features of using the logistic regression method to predict employees churn. [14]. This technique is one way to answer questions like "why do people leave a company?", "how can we influence turnover rates?", and "who is most likely to leave in the next year?" The goal of any type of regression is to predict an outcome using some factors. To predict churn (the dependent variable) employee data (the independent variables) is use. Logistic regression, simply put, is a form of regression that is used when the predicted outcome is 1 or 0. This is exactly what is needed to predict turnover: people either leave or stay with the company.

Logistic regression determines the probability of an event occurring based on the set of attribute values *y*. For this purpose, the so-called dependent variable y is introduced, which can take only one of two values. As a rule, these are the numbers 0 (event did not happen) and 1 (event did happen), and a set of independent variables (also called attributes, predictors or regressors) that are real values $x_1, x_2, \ldots, x_n$. They are based on the values needed to calculate the probability of assuming a certain value of the dependent variable. Let's introduce a fictitious attribute $x_0 = 1$ to simplify the writing later.

It is assumed that the probability of the event occurring $y = 1$:

$$P\{y = 1|x\} = f(z),$$

where $z = \theta^T x = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$, where x and θ are column vectors of values of the independent variable $1, x_1, \ldots, x_n$ and parameters (regression coefficients) are real numbers θ_0, ..., , and $f(z)$ is the so-called logistic function (sometimes also called sigmoid or logit function):

$$f(z) = \frac{1}{1 + e^{-z}}.$$

Since y takes only the values 0 and 1, the probability of taking the value 0 is:

$$P\{y = 0|x\} = 1 - f(z) = 1 - f\left(\theta^T x\right).$$

In general, the distribution function *y* for a given *x* can be written as:

$$P\{y|x\} = f\left(\theta^T x\right)^y \left(1 - f\left(\theta^T x\right)\right)^{1-y}, y \in \{0,1\}.$$

2. Decision tree.

Decision tree methods are a family of algorithms based on creating a hierarchical structure based on the answer to a set of "Yes" or "No" questions. The create of decision tree methods for data analysis is associated with the C4.5 and C5.0 algorithms, the CART (Classification and Regression Trees) algorithm and the random forest method [14].

The basis of most popular decision tree training algorithms is the principle of data partitioning. This principle is implemented by the following algorithm. Consider a training set S containing n examples, each of which is given a class label $Ci(i = 1..k)$ and n attributes $Aj(j = 1..m)$, which are supposed to determine whether an object belongs to a particular class. Then three cases are possible:

1. All examples of the set S have the same class label Ci, that is, all training examples belong to only one class. Obviously, training in this case does not make sense, since all examples the presented models will be of the same class, which will "learn" to recognize the model. The decision tree itself in this case will be a leaf associated with the class Ci. The practical use of such a tree is meaningless, since it will attribute any new object only to this class.

2. The set S does not contain any examples at all, i.e. is an empty set. In this case, a leaf will also be created for it (applying the rule to create a node to an empty set is meaningless), the class of which will be selected from another set (for example, the class that occurs most often in the parent set).

3. The set S contains training examples of all classes Ck. In this case, it is necessary to divide the set S into subsets associated with classes. To do this, one of the attributes Aj of the set S is selected, which contains two or more unique values $(a1, a2, \ldots, ap)$, where p is the number of unique values of the attribute. Then the set S is partitioned into p subsets $(S1, S2, \ldots, Sp)$, each of which includes examples containing the corresponding attribute value. Then the next attribute is selected and the partitioning is repeated. This procedure is repeated recursively until all examples in the resulting subsets are of the same class.

In the process of constructing a decision tree, it is necessary to select a partitioning attribute. The best partition attribute will be the one that provides the maximum information gain of the resulting node relative to the original:

$$Gain(A) = Info(S) - Info(S\_A),$$

where $Info(S)$ is the information associated with the subset S before partitioning,

$Info(S\_A)$ is the information associated with the subset obtained by partitioning by attribute A.

Thus, the task of choosing a partition attribute in a node is to maximize the value of $Gain(A)$, which is called the information gain.

Experimental part. For the practical implementation of predictive methods (the decisions tree and logistic regression) we will use the modern package of IBM SPSS Modeler application programs [9]. Modeling employee churn is a description and analysis to predict the category to which a particular employee belongs. In our case, the simulation of employee churn is based on signs of churn: work (True), left (False). We estimate churn using the constructed flow and the company's personnel database. First, we estimate the importance of the predictor, our model determined that the most important for forecasting is work experience at the start of cooperation, the value of which = 0.15 (Fig. 3).

After completing the construction of the flow, the final result was the model of building a decision tree, which is presented on the Fig. 2.

Using a decision tree and analyzing data on an individual worker, it is possible to simulate the situations under which he will stay in the company, and under which he will leave the company.

The CART algorithm has two children at each node of the decision tree. A rule is formed at each node. It allows you to divide the set of examples of a node (the training set) into two parts. The first part contains examples in which the rule is fulfilled (usually the child is on the right), and the second part in which the rule is not fulfilled (the child is on the left). The partition quality evaluation function is used to determine the optimal rules. For example, based on decision trees, it can be assumed that workers who have had overtime, the number of companies they have worked for >2, will leave the company with a probability of 100%, instead, workers, who worked for our company in the same company, or had no work experience at all, will remain. This may be due to the fact that a worker who already has experience working in another company/companies without overtime with a standardized work schedule, while working in a new com-

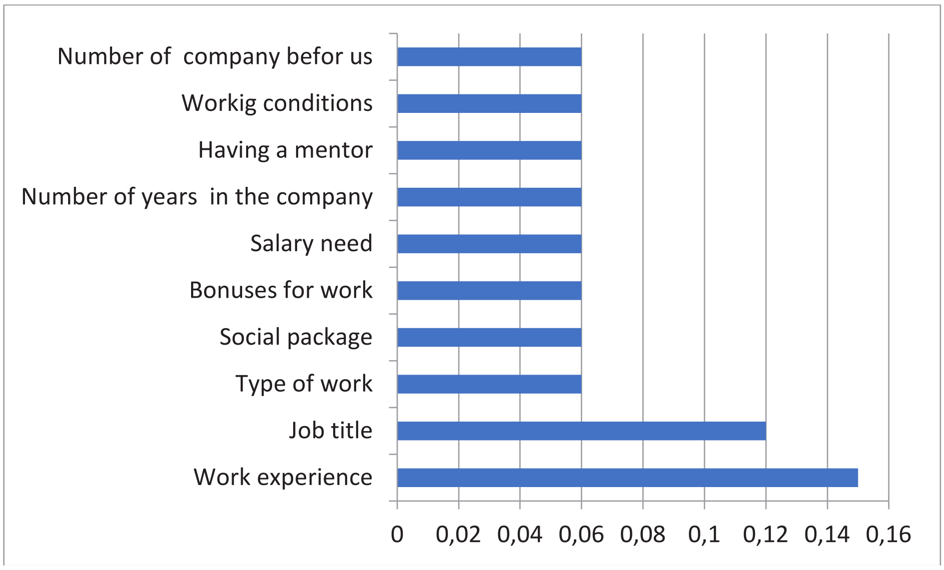

Fig. 1. Estimate the importance of the predictor

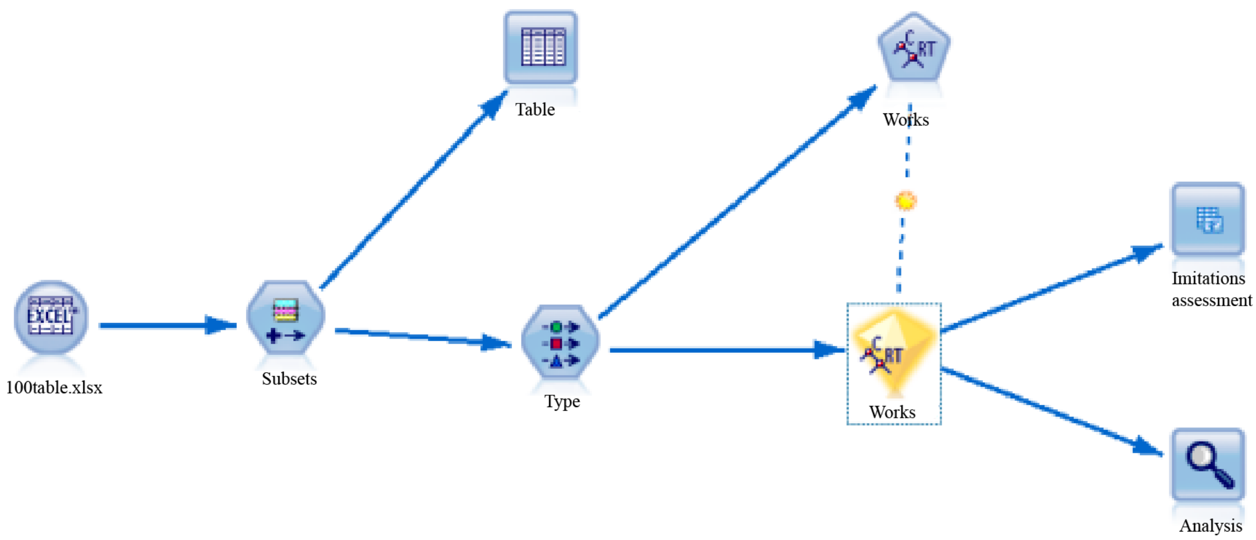

Fig. 2. Decision tree construction flow

pany with overtime, can compare the conditions and change the place of work.

Let's consider another branch of our tree. An employee who did not have overtime may leave because he does not have a social package, and for him this is more important than anything.

Let's also consider the option of a worker who does not have overtime. According to the analyzed data, a person who has work experience <=2.8 years with a 100% probability will remain working for the company. An employee who has work experience in the company >2.8 years, work experience in our company <3.7 years is highly likely to leave the company.

Therefore, it can be concluded that the main factors that affect the churn of employee are: the presence of overtime, the position they hold, the social package, work experience and the number of companies in which the employee worked before starting work in our company.

The goal of the whole process of building a decision tree is to create a model that can be used to classify cases and decide what values the objective function can take, given several variables as inputs. Thus, the classification of each new case occurs when moving down to the letter, which will tell us the value of the objective function in each specific case.

In our case, when it is necessary to make a decision about the worker (the objective function can take the values "working" and "left") based on information about the worker (for example: work experience at the time of the start of cooperation, availability of overtime, marital status, and so on).

To build a model using logistic regression, we perform the same settings as for the decision tree, namely adding nodes: source, subset, table and types. As a result of construction, we get a logistic regression model, which is presented on the Fig. 3.

After running the logistic regression model, a classification table was obtained, which shows how correctly the built model predicts (Table 3). A total of 100 records were processed. The percentage of correct predictions for employees with the attribute "work" is 94.2%, and among those who "left" – 83.3%. The average percentage is 89.0%.

Table 3

Classification table

Sign	Predicted		
	Works	Left	Percentage of correct
Works	49	3	94,2
Left	8	40	83,3
The overall percentage			89,0

For employed persons, 49 correct predictions were obtained, and 40 correct predictions for persons who are not employed. When training the model, 8 types of false classifications for those who are employed, and 3 false negatives for those who are not employed, must be omitted.

Next, we will analyze the coefficients of the so-called chances, which show the statistical

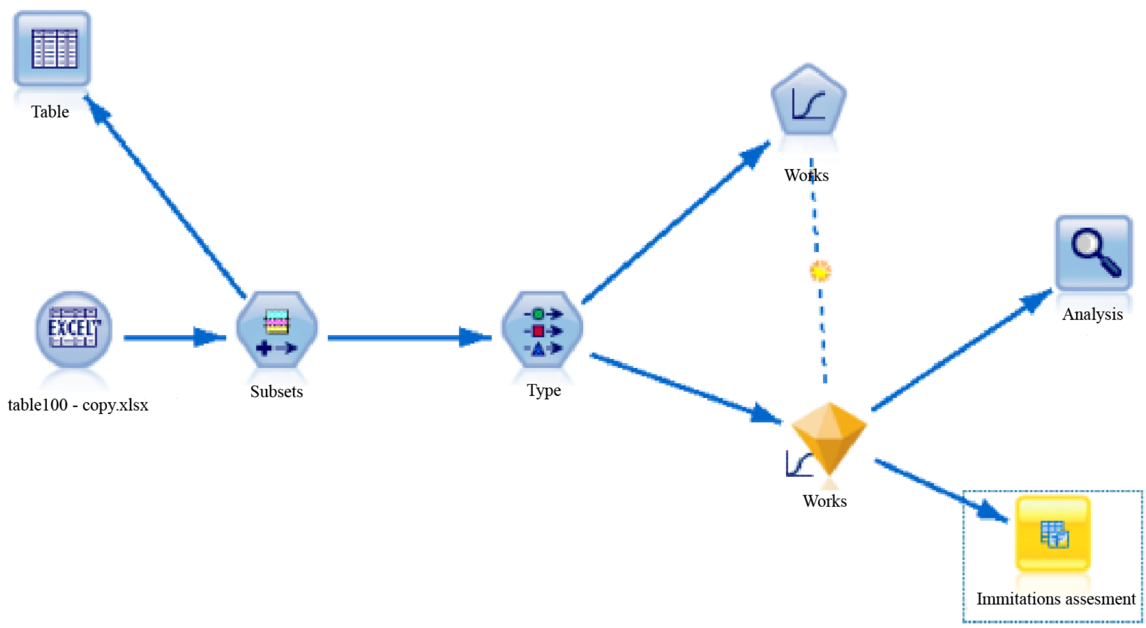

Fig. 3. Logistic regression model

influence of variables on the churn of employee (Fig. 4). As we can see, these are the variables included in the model. In this table, it is worth paying attention to the evaluations of the variables. When we build a model based on these estimates, these are the odds ratios. The table shows the confidence interval for the odds ratios. Indicators in column B are the actual beta coefficients on which the model is built. Odds ratios determine the influence of variables, the higher this ratio (greater than 1), the more likely this variable will affect the dismissal of the worker or his further work in the company.

Analysis of the simulation results shows that the significance level for the model is < 0.05, which means that we have a significant model and will predict well.

The analysis of the simulation results also shows how many variants of the dependent variable are explained by the predictors. In the built model, this is about 73% of the result, or 73 percent of the options in the result will be predicted by our predictors (or affected by our predictor). Thus, it can be concluded that a significant percentage of the results will be predicted by the previously selected predictors.

Therefore, it can be concluded that the result was correctly predicted by the model for 89%. This is much better than the null hypothesis before model training, as it was only 52%.

Comparing the obtained results and works based on the comparison tables, where the number and percentage of correct and incorrect, trained and tested stream entries are presented, the decision tree model showed results of 91.3% and 74.19% correct for trained and tested, respectively, and the logistic regression model – 88.41% and 90.32%. Thus for the decision tree there were better results for the trained set, and for regression the tested set.

Also evaluating the performance by the indicators trained in the decision tree 0.514 and 0.698 respectively (True, False) and logistic regression 0.463 and 0.691 respectively. According to the indicators of tested decision trees 0.458 and 0.333 respectively (True, False) and logistic regression 0.601 and 0.601 respectively.

To obtain the general results of the simulation for two models, it is necessary to run the model analysis node for execution. This node makes it possible to assess the adequacy of the constructed model. After its execution, the system displays the results of the model evaluation for the decision tree (Fig. 5).

Let us consider the values of the model evaluation indicators, namely the values of the AUC indicator and the Gina coefficient. AUC is area under curve ROC (receiver operating characteristic) and the Gina coefficient is quantitative indicator that shows the degree of inequality of the distribution options. The values of these indicators are always

Variables in the equation

	B	Root mean squared error	Wald	d.f.	Value	Exp (B)
Work experience by the employment(in years)	,948	,359	6,973	1	,948	2,580
How many companies have worked before us	-,074	,471	0,025	1	-,074	,929
Years in company	,597	,427	1,953	1	,597	1,816
Education	6,178	3,530	3,063	1	6,178	481,843
Social package	-26,258	40193,018	,000	1	-26,258	,000
Working conditions	-502	724	482	1	-502	605

Fig. 4. Statistical characteristics of model variables

'Subsets'	1_Training		2_Testing	
Model	AUC	Jeanie	AUC	Jeanie
$R-Works	0,951	0,902	0,817	0,634

Fig. 5. Results of decision tree model analysis

'Subsets'	1_Training		2_Testing	
Model	AUC	Jeanie	AUC	Jeanie
$L-Works		0,881	0,908	0,815

Fig. 6. The results of the analysis by logistic regression

between 1 and 0, the larger the value, the better the classifier.

Next, we perform regression analysis. The results of the logistic regression analysis are presented in Fig. 6.

The AUC and Gina scores are almost the same, but the decision tree is several hundredths better in terms of AUC than logistic regression with 0.951 and 0.817, and logistic regression with 0.941 and 0.815. The Gina coefficient also shows that the decision tree is better, 0.902 and 0.634, respectively, and the logistic regression is 0.881 and 0.815, respectively.

Thus, both models showed satisfactory results. But according to the AUC and Gina indicators, the preference can be given to the decision tree.

Conclusions. Thus, the study substantiated the approach to employee turnover management using predictive analytics methods. The theoretical foundations of employee turnover in a company are the types of employee turnover, such as natural turnover and intra-organizational turnover. An important element of employee turnover management are the components of its measurement, namely, the overall level of turnover and its comparison with turnover in other companies on the market (benchmarking); the dynamics of the turnover coefficient over time; changes in the structure of the causes of turnover; changes in the turnover of different categories of personnel. In addition, it is necessary to take into account the factors that influence employee turnover and their regulation: material, organizational, interpersonal.

For the practical implementation of employee turnover management, the use of mathematical methods and information technologies was proposed, which allowed building and analyzing predictive models of employee turnover.

Before modeling, the choice of indicators that influence employee turnover was justified. Next, flows were constructed for modeling the outflow using the decision tree and logistic regression methods. The experimental results showed that the accuracy of the predictive model by both methods is quite high. Having evaluated the data obtained on the analysis of the constructed models by the decision tree and logistic regression methods, a conclusion was made about the adequacy of both models.

Comparing the results of the work, which presents the number and percentage of correct and incorrect trained and tested records in the flow, the decision tree model showed results of 91.3% and 74.19% correct for trained and tested, respectively, and the logistic regression model – 88.41% and 90.32%. These results indicate a fairly high accuracy of the forecast

The scientific significance of the provisions and recommendations developed in the article is that they allow solving the scientific and practical task of developing methodological foundations and practical recommendations for the use of computer modeling methods in predicting the outflow of enterprise personnel.

The validity of the scientific results of the study is determined by a multi-faceted consideration of the problem, research and experimental confirmation of the adequacy of the developed predictive model.

The practical significance of the results obtained lies in the development of recommendations for the construction and analysis of models for predicting employee turnover for the enterprise.

Bibliography:

1. Catlett F. E. Management Strategies to Improve Employee Retention and Organizational Performance. Dissertations of Doctor of Business Administration Walden*:* 2024, 136 p. URL: https://scholarworks.waldenu.edu/dissertations/15867

2. Paruzel A., Schmidt L., Maier G. W. Corporate social responsibility and employee innovative behaviors: A meta-analysis. *Journal of Cleaner Production.* 2023. № 393. URL: https://doi.org/10.1016/j.jclepro.2023.136189

3. Starkova O., Bondarenko D., Hrabovskyi Y. Providing software support for economic analysis. *Technology Audit and Production Reserves.* 2023. № 5/2 (73). P. 34–39. URL: http://repository.hneu.edu.ua/handle/123456789/30456

4. Ushakova I., Skorin Y., Shcherbakov A. Methods of quality assurance of software development based on a systems approach. *III International Scientific And Practical Conference “Information*

Security And Information Technologies", September 13–19, 2021, Odesa, Ukraine. P. 158–168. URL: https://ceur-ws.org/Vol-3200/paper22.pdf

5. Ushakova I., Hrabovskyi Ye. Methodology for developing an information site with Workflow support for publishing articles. *Development Management.* 2022. Vol. 21. No. 3. P. 20–28. URL: https://doi.org/10.57111/devt.20(3).2022.20-28.

6. Hrabovskyi Ye., Bondarenko D., Kobzev I. Improving the technology for constructing a software tool to determine the similarity of raster graphic images. *Eastern-European Journal of Enterprise Technologies.* 2024. № 1/2(127). P. 16–35. URL: https://journals.uran.ua/eejet/article/view/298744/291778

7. Hrabovskyi Y., Bondarenko D., Starkova O. Development of a methodology for creating a website for online healthcare services. *EUREKA: Physics and Engineering.* 2024. № 5. P. 183–200. URL: https://journal.eu-jr.eu/engineering/article/view/3440/2496.

8. Erden S. B., Erişen M., Sarıgül Y. N., Eken B., Özcan T. HR Analytics in Retail: Predicting Employee Churn with Machine Learning. *Intelligent Industrial Informatics and Efficient Networks Proceedings of the INFUS 2024 Conference.* 2024. Vol. 3. P. 109–116. URL: https://link.springer.com/book/10.1007/978-3-031-67192-0

9. Meddeb E. The Human Resource Management challenge of predicting employee turnover using machine learning and system dynamics. *CEUR-WS.* 2021. Vol. 2991. P. 184–196. URL: https://ceur-ws.org/Vol-2991/paper16.pdf

10. Jiaxuan Zh., Zinan C., Haichen Q., Meng W. Management and prediction of employee turnover in enterprises based on big data analytics and machine learning. *Proceedings of the 2nd International Conference on Software Engineering and Machine Learning.* 2024. P. 103–108. URL: http://dx.doi.org/10.54254/2755-2721/76/20240573

11. Ushakova I., Hrabovskyi Ye. Modeling and Selection of a Distance Learning System for a Higher Education Institution Based on the Method of Hierarchy Analysis Using the DSS. Вчені записки ТНУ імені *В.І.* Вернадського. 2023. №. 34 (73). P. 246–254. URL: https://doi.org/10.32782/2663-5941/2023.2.1/39

12. Bandyopadhyay N., Jadhav A. Churn Prediction of Employees Using Machine Learning Techniques. *TEHNIČKI GLASNIK*. 2021. Vol. 15. № 1. P. 51–59. URL: https://doi.org/10.31803/tg-20210204181812

13. Ushakova I. Simulation of the impact social media on promoting education services. Proceedings of the 12th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Kyiv, Ukraine, June 21–24, 2016. Vol. 1614. P. 575–590. URL: https://ceur-ws.org/Vol-1614/paper_60.pdf

14. Ushakova I., Dorohov O., Dorohova L., Malyarets L. Customer churn predictive modeling by classification method. *Bulletin of the Transilvania University of Brasov. Series III: Mathematics, Informatics, Physics.* 2020. Vol. 13 (62). № 1. P. 347–362. URL: https://doi.org/10.31926/but.mif.2020.13.62.1.26

15. Ushakova I., Hrabovskyi Ye., Szymczyk K. Analytical Decision List Algorithm for Managing Customer Reaction to a Marketing Campaign. *Polish Journal of Management Studies.* 2023. Vol. 28. № 1. P. 360–379. URL: https://pjms.zim.pcz.pl/article/541885/en

Ушакова І.О., Бондаренко Д.О., Чирва Ю.Є., Знахур Л.В. КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ ВІДТОКУ СПІВРОБІТНИКІВ КОМПАНІЇ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА ПРОГНОЗНОЇ АНАЛІТИКИ

У дослідженні запропоновано підхід до управління відтоком співробітників за допомогою методів машинного навчання, прогнозної аналітики та за підтримки інформаційних технологій. В рамках досліджування було визначено теоретичні основи плинності кадрів, проведено її класифікацію за різними типами та визначено ключові компоненти для її вимірювання. Крім того були обґрунтовані основні фактори, що впливають на показники відтоку кадрів та їх потенційне регулювання. Дискурс підкреслює важливість стратегій утримання співробітників, наголошуючи на таких елементах, як винагорода, планування роботи, оцінка ефективності, програми навчання та можливості кар'єрного зростання. Автори пропонують використовувати методології дерева рішень і логістичної регресії для прогнозування відтоку співробітників, вибираючи двійковий критерій класифікації, який розрізняє співробітників, які залишаються в компанії, і тих, хто йде. Розроблено дві прогностичні моделі, які демонструють значні показники точності: модель дерева рішень демонструє вражаючу точність 91,3% щодо даних навчання та 74,19% даних тестування, тоді як модель логістичної регресії вказує на точність 88,41% даних навчання та 90,32% даних тестування. Отримані результати підкреслюють надійність запропонованих моделей у прогнозуванні відтоку кадрів.

Крім того, у статті описано практичне застосування цих моделей, що надає важливу інформацію для організацій, які прагнуть реалізувати керовані даними стратегії в напрямку утримання співробітників в компанії. Завдяки застосуванню передових аналітичних методів дослідження *створює цінні методології та рамки для компаній, які борються з проблемами відтоку працівників, позиціонуючи себе як найважливіший ресурс для спеціалістів із відділу людських ресурсів та керівників організацій, які прагнуть підвищити стабільність та продуктивність співробітників. Загалом дослідження підкреслює потенціал застосування методів машинного навчання, прогнозної аналітики та комп'ютерного моделювання для прийняття обґрунтованих рішень щодо утримання співробітників і підтримки здорового клімату в компанії.*

Ключові слова: *комп'ютерне моделювання, машинне навчання, логістична регресія, дерево рішень, відтік співробітників, прогнозна аналітика.*