IRSTI 50.43.19 https://doi.org/10.26577/JPCSIT.2023.v1.i2.08

M. Zhanuzakov*, G.T Balakayeva

Al-Farabi Kazakh National University, Kazakhstan, Almaty
*e-mail: zhanmuha01@gmail.com

DIGITALIZATION OF ENTERPRISE HUMAN-RESOURCE MANAGEMENT USING MACHINE LEARNING MODELS

Abstract. Improving enterprise`s efficiency is crucial in today's world. Thus, assessing employees' contributions to the enterprises is essential. As a result, staff competencies are the primary focus in large enterprises. Many enterprises require human resource management in order to effectively analyze it. In this study, the authors conducted research and developed a model for assessing and analysing personnel utilizing software and artificial intelligence. During the research, the digitalization of the following services was carried out: surveys, feedbacks and predictor tools. Employee ratings are gathered through surveys, and estimations are created using both positive and negative comments from coworkers. This article describes the development of a machine learning model for predicting employee attrition. During which, various machine learning algorithms were evaluated, with KNN and Decision tree classifier producing the most promising results in terms of accuracy, precision, recall, and F1 score. The article also provides a data collection for storing employee ratings using MongoDB.

Key words: digitalization, enterprises, rating, machine learning, model, evaluation, assessment, employee.

Introduction

Digitalization refers to the use of digital tools to transform traditional business processes and practices into digital ones. This includes the use of computers, software, and the internet to automate and streamline business operations. One area of business that has seen significant digitalization in recent years is human resource management (HRM) [1].

HRM involves all the activities and processes related to managing and developing an organization's workforce. This includes recruiting and hiring employees, training and development, performance management, and payroll and benefits administration [2]. Traditional HRM practices often rely on manual processes, such as paper-based forms and manual data entry. However, with the rise of digital technologies, many organizations are now looking to digitalize their HRM to improve efficiency and effectiveness.

One way that organizations are using digital technologies to transform HRM is by implementing machine learning (ML) models. ML as a technology of artificial intelligence, allows computers to learn and improve their performance without being explicitly programmed. In the context of HRM, ML models can be used to automate various HR tasks, such as resume screening, employee performance evaluation, and succession planning.

There have been a few investigations on this topic [3–7]. These articles cover a wide range of topics related to the use of machine learning in the digitization of business human-resource management activities, such as recruiting and retention, employee performance, and skill evaluation. They demonstrate several applications and case examples, as well as give insights into the possible benefits and obstacles of adopting machine learning models in human resource management. We analyzed the works of these authors and developed new ways of digitalization of the HRM.

Methods and models

As a measure value of employee`s competencies, we used rating. A rating is a numerical value that represents the employee's current level of competence. It ranges from 0 to 5. The rating is a float (approximation to 2 decimal places).

We purpose a rating analyses model which makes analyses on the rating of employees each month using the following methods and approaches:

1. Surveys
2. Feedbacks from colleagues
3. Attrition prediction model using machine learning algorithms

4. Promotion prediction model using machine learning model

Surveys**.** A survey is a collection of pre-written questions produced and revised by business management. At the conclusion of each month, all enterprise employees take part in a survey in which they rate their coworkers.

Surveys are entirely optional. Employees that are compelled to do surveys will produce skewed data since they will often offer unrealistic scores in order to complete the survey procedure as fast as possible.

The information gathered from these surveys will be utilized to develop estimates for the next months. As a result, it is critical that survey data be as "clean" as feasible.

Reviews. Reviews are text messages from coworkers that provide feedback on an employee. There are two types of feedbacks:

1) Positive (complement)
2) Negative (complaint).

They are only viewed by corporate managers for psychological reasons, as negative feedback might damage employee motivation and mental health.

Anonymous feedback is possible. When negative input is generated, the current month rating is reduced by 10%; when good feedback is generated, it is increased by 15%. Because favorable feedback is uncommon and often difficult to obtain.

If the user's current rating is "null" (not yet evaluated), negative feedback will result in a rating of 4.5 (10% of 5), whereas good feedback will result in a rating of 5.

The comments might be of the "compliment" or "complaint" variety. The number of compliments and complaints received is also displayed, allowing HR managers to rapidly identify unprofessional and positive employee qualities (Figure 1).

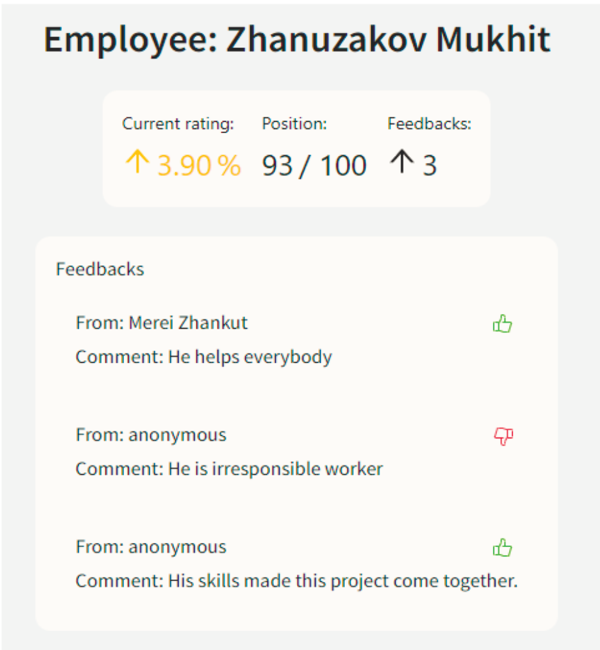

Figure 1 – Employee page with ratings and feedbacks

Employe attrition prediction model. Employee attrition refers to the natural process of employees leaving an organization over time, either voluntarily or involuntarily, and not being replaced. Thus, for HR managers, it is essential to have tools to predict early turnover of their employees to start looking for replacements. For this reason, we have developed an attrition prediction model using machine learning algorithms.

Dataset. IBM data of employees was used as a dataset for training the prediction model. This dataset was created by IBM data scientists to investigate the factors that contribute to employee attrition and answer critical inquiries [8].

The dataset has 35 features and 1470 samples. The dataset does not have any duplicate or missing values.

The key feature is the attrition result, which identifies whether the employee left the enterprise or not (yes/no). The percentage of attrition can be viewed in Figure 2.

```
No     83.877551
Yes    16.122449
Name: Attrition, dtype: float64
```

Figure 2 – Attrition percentage

Data processing. The "Over18", "StandardHours" features were dropped, as they provide constant values and do not have any meaningful impact on predicting employee attrition.

All categorical features such as "Attrition", "OverTime", "Gender", "BusinessTravel", "Department", "EducationField", "JobRole", "MaritalStatus" have been converted to unique numeric values.

Using correlations between independent variables, it has been identified that "EmployeeCount" has not relation to "Attrition" (see Figure 3), thus was dropped from the table. Whereas "TotalWorkingYears", "JobLevel" and "MonthlyIncome" had the highest correlations.

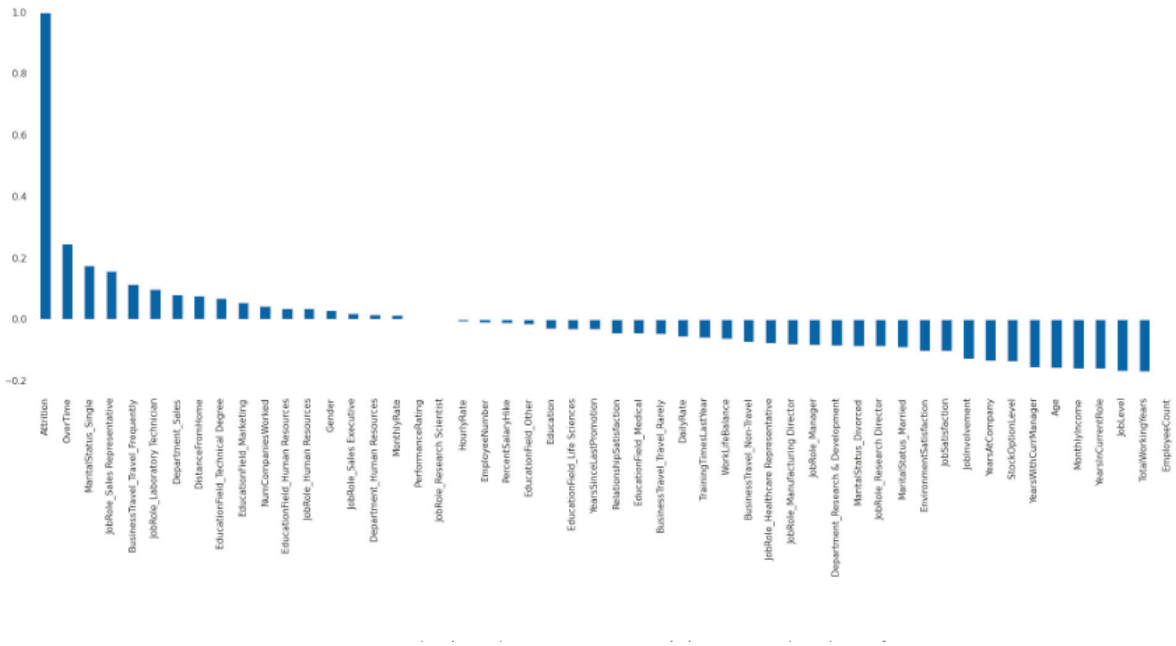

Figure 3 – Correlation between "attrition" and other features

Building model. To train the dataset, the data was split into three collections: 80% – training set, 20% – test set. In order to assess overfitting and underfitting, the test data was further split into validation (50%) and test sets (50%).

To select the machine learning algorithm, the following classifiers were trained and tested using the dataset:

1. Decision tree classifier
2. Random forest classifier

3. Support Vector Machines (SVM) classification
4. K-Nearest Neighbors (KNN)
5. Gradient Boosting

To identify the best value for parameter "k" (number of neighbors), we used cross validation technique for each k starting from 1 to 31. The best accuracy was produced when k was 1 (87%), then the accuracy started to incline. The overall results are displayed in Figure 4.

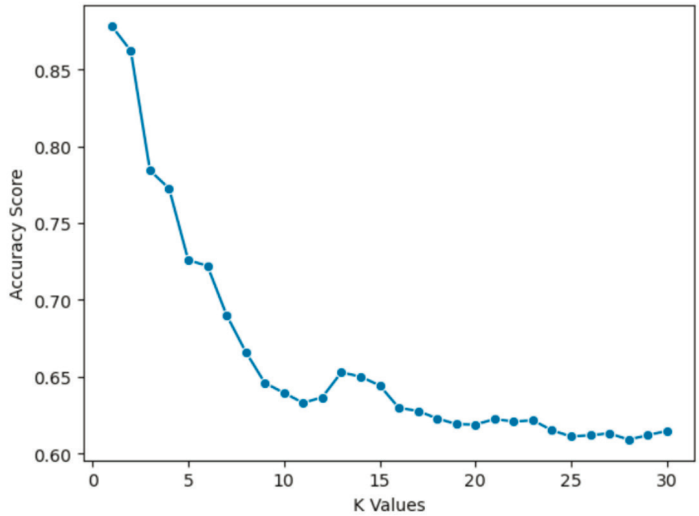

Figure 4 – Using Cross Validation to Get the Best Value of k

To build the models, scikit-learn python library was used.

Results and discussion. To assess the performance of the developed model, the following metrics were used: accuracy, precision, recall and f1 score. The trained model was evaluated using both validation and test sets, providing more accurate performance results. The results of these tests are presented in Table 1.

Table 1 – Results of performances of the trained models

Model	Accuracy (%)		Precision (%)		Recall (%)		F1 (%)	
	V	T	V	T	V	T	V	T
Decision tree classifier	89	95	87	93	95	98	90	95
Random Forest classifier	83	87	86	92	81	81	83	86
SVM	88	84	88	84	89	86	89	85
KNN	93	90	93	86	94	97	93	91
Gradient Boosting	82	83	84	83	81	86	82	84

* V – result on validation set
* T – result on test set

During the tests, KNN and Decision tree classifier has showed the best results.

The priority metric for the tests is F1, as it provides an assessment of both precision and recall.

In these terms, the best-performing model is KNN. Even though it showed a lower F1 value on the test set (91%) compared to the Decision Tree Classifier (95%), its results are more consistent, meaning that the difference between the results on different sets is smaller.

Our model demonstrated better performance in F1 score compared to the model presented in the related work [10]. Specifically, our KNN model achieved a higher score, with their model achieving an F1 score of 88%.

Comparing to the work in [11] that uses the same dataset, KNN achieved 93% F1 score with K = 3 using adaptive synthetic approach. Our model was able to show close results by picking the right k and using only correct features.

3 Application and usage of machine learning models

We have developed two machine learning models:

1) Employee attrition prediction model.

2) Employee promotion prediction model. We have developed and described promotion prediction model using machine learning methods in the paper [8]. In this work, we only describe its usage.

A Python web server was designed that uses the developed model for forecasting promotion and attrition of the selected employee. Web server takes employee ratings and information for the current month for each competency from the database, feeds the data into the prediction models, then returns the models` results to the client application (Figure 5).

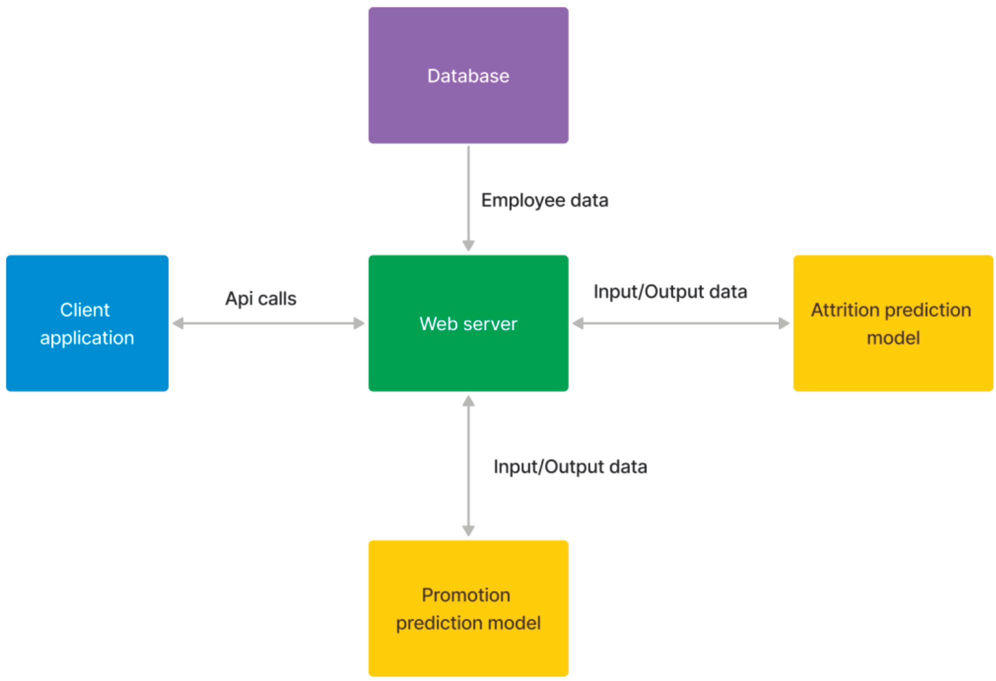

Figure 5 – Prediction models usage

The developed attrition prediction model helps managers to report early attrition of the employee. It also aids when deciding to promote employees to the next position alongside with promotion prediction model. When promotion prediction model predicts that employee should be promoted, it is also possible that this employee can leave the company soon. Thus, managers should consider the results of the attrition prediction model, before making any decision.

Creation of a data collection in MongoDB

MongoDB was used as a database for storing ratings and employees` data. MongoDB is a popular NoSQL database that is often used for storing big data collections. MongoDB is a powerful and flexible database that is well-suited for handling big data collections. Its scalability, flexibility, performance, availability, and analytics capabilities make it a great choice for organizations that need to store and process large amounts of data [12 – 14].

"Ratings" data collection has been created to store employee`s rating data. Rating collection object properties have the following properties:

```
{
_id: String,
assessedUserId: String,
assessorUserId: String,
competencyId: String,
createdAt: Date,
deletedAt: Date,
value: Number
}
```

Whenever the employee gets the promotion or leaves the enterprise, its current rating and state is stored in the database with the following information:

```
{
    _id: String,
    employeeId: String,
    currentRating: Number,
    createdAt: String,
    isPromoted: Boolean,
    isAttrition: Boolean
}
```

This data then can be further used for building datasets for other ML models in the future research.

Acknowledgments

The Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan financed this research (Grant #AP09259208).

Conclusion

The key issue for modern enterprises is determining the quality of their employees' productivity and professional development. The created methodology and tools for assessing and predicting corporate personnel were discussed in this study. The aim of the research was to digitalize human resource management activities and build prediction tools. This was achieved using software development and machine learning techniques. The used software technologies: Python, scikit-learn, MongoDB.

A model for predicting employee attrition was developed and used for HRM alongside employee promotion model. In order to train the employee attrition prediction model, the dataset was split into three collections for training, testing and validation purposes. Several machine learning algorithms were trained and tested, with KNN and Decision tree classifier showing the best results in terms of accuracy, precision, recall and F1 score. KNN was found to be the best-performing model due to its consistent results across different sets. Overall, the selection of the right value for k and features led to close results with significantly reduced computational costs.

The following activities were digitalized during the work: surveys, feedbacks, rating reports and prediction tools.

References

1. Vardarlier, P. "Digital Transformation of Human Resource Management: Digital Applications and Strategic Tools in HRM." In Digital Business Strategies in Blockchain Ecosystems, edited by U. Hacioglu, 215-232. Springer, 2020. https://doi.org/10.1007/978-3-030-29739-8_11.
2. Baykal, Elif. "Digitalization of Human Resources: e-HR." In Handbook of Research on Digital Transformation and Innovation in the Global Economy, edited by Y. K. Dwivedi and N. Rana, 197-213. IGI Global, 2020. doi: 10.4018/978-1-7998-0035-4.ch013.
3. Cao, L., Zhang, Y., & Wei, J. (2020). A machine learning-based human resource management system for enterprises. Future Generation Computer Systems, 107, 1070-1079.
4. Alghamdi, M., Alqahtani, R., Alruwaili, M., Alghamdi, N., & Alhajri, H. (2021). Digitalization of human resource management: A machine learning approach for predicting employee retention. IEEE Access, 9, 70387-70399.
5. Shih, M. S., Cheng, Y. H., & Chang, Y. C. (2019). A machine learning-based human resource management system for employee recruitment and retention. International Journal of Information Management, 47, 310-320.
6. Saha, S., & Kumar, V. (2019). A machine learning-based human resource management system for predicting employee performance. Journal of Business Research, 96, 313-323.
7. Chen, Y., Yang, Y., & Li, X. (2021). A machine learning-based human resource management system for employee skill assessment. Journal of Intelligent Manufacturing, 32, 1791-1802.
6. Tabular summary of HR analytics dataset. Url: https://zenodo.org/record/4088439#.ZFSToHZByUl, visit date: 05.05.2023.

7. Zhanuzakov, M. and Balakaeva, G. "Prediction of Employee Promotion Based on Ratings Using Machine-Learning Algorithms." Journal of Physics & Mathematical Sciences 77, no. 1 (March 2022): 106-111. https://doi.org/10.51889/2022-1.1728-7901.14.

Yedida, R., Reddy, R., Vahi, R., Jana, R., GV, A., & Kulkarni, D. (2018). Employee attrition prediction. arXiv preprint arXiv:1806.10480.

8. S. S. Alduayj and K. Rajpoot, "Predicting Employee Attrition using Machine Learning," 2018 International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 2018, pp. 93-98, doi: 10.1109/INNOVATIONS.2018.8605976.

9. Balakayeva, G., and D. Darkenbayev. "The Solution to the Problem of Processing Big Data Using the Example of Assessing the Solvency of Borrowers." Journal of Theoretical and Applied Information Technology 98, no. 13 (2020): 2659-2670. ISSN 1992-8645.

10. Balakayeva G.,C. Phillips ,D.K. Darkenbayev , M. Turdaliyev Using Nosql For Processing Unstructured Big Data News Of The National Academy Of Sciences Of The Republic Of Kazakhstan, Series Of Geology And Technical Sciences.- 2019 – Volume 6, Number 438 – P.12-21.

11. Balakayeva, G.T., Ezhichelvan, P., Tursynkozha, M.K. Analysis, Research and Development of an Innovative Enterprise Digitalization System for Remote Work, International Journal of Mathematics and Physics, – 2022. – P. 19-29