МРНТИ 20.53.21
УДК 004.67

https://doi.org/10.51889/2022-1.1728-7901.14

PREDICTION OF EMPLOYEE PROMOTION BASED ON RATINGS USING MACHINE-LEARNING ALGORITHMS

Zhanuzakov M.B.[1], Balakaeva G.T.[1]*

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

Abstract

Employee promotion is an important aspect in human-resource management process. Thus, it is crucial to correctly decide, whether an employee should or should not be promoted based on its current and past ratings. For that purpose, the research has been carried out to develop employee`s promotion prediction models by using different machine-learning classification algorithms. In this research, experiments on simulated dataset was performed using Gradient Boosting Classifier, Random Forest Classifier and Keras Neural Network. Through a complex assessment process, the performance of these supervised machine learning algorithms for predicting employee advancement was analyzed using assessment metrics. Our study uses simulated employee data as the training dataset, and we developed a web application for our study to display forecast results on new inputs.

Keywords: machine learning, rating, neural networks, deep learning, classification, forecasting, personnel management, promotion.

Аннотация

М.Б. Жанузаков[1], Г.Т. Балакаева[1]
[1]Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан

ПРОГНОЗИРОВАНИЕ ПОВЫШЕНИЕ СОТРУДНИКА ПО СЛУЖБЕ НА ОСНОВЕ РЕЙТИНГА СОТРУДНИКА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

Продвижение сотрудников, важный аспект процесса управления человеческими ресурсами. Важно правильно решить, следует ли продвигать сотрудника по карьерной лестнице, исходя из его нынешних и прошлых рейтингов. С этой целью было проведено исследование по разработке моделей прогнозирования продвижения сотрудников с использованием различных алгоритмов классификации машинного обучения. В исследовании эксперименты на смоделированном наборе данных были выполнены с использованием классификатора градиентного усиления, классификатора случайного леса и нейронной сети Keras. Посредством сложного процесса оценки, эффективность этих контролируемых алгоритмов машинного обучения для прогнозирования продвижения сотрудников по службе была проанализирована с использованием метрик оценки алгоритмов. В качестве набора данных для обучения используются смоделированные данные сотрудников. В ходе исследования разработано веб-приложение для отображения результатов прогнозов моделей.

Ключевые слова: машинное обучение, рейтинг, нейронные сети, глубокое обучение, классификация, прогнозирование, управление персоналом, продвижение.

Аңдатпа

М.Б. Жанузаков[1], Г.Т. Балакаева[1]
[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ПАЙДАЛАНУ АРҚЫЛЫ ҚЫЗМЕТКЕРДІҢ РЕЙТИНГІ НЕГІЗІНДЕ ҚЫЗМЕТКЕРДІҢ ҚЫЗМЕТ БОЙЫНША ЖОҒАРЫЛУЫН БОЛЖАУ

Қызметкерлерді жоғарылату адам ресурстарын басқару процесінің маңызды аспектісі болып табылады. Сондықтан қызметкердің қазіргі және бұрынғы рейтингіне сәкес қызметкерді жоғарылату туралы дұрыс шешім қабылдау өте маңызды. Осы мақсатта әртүрлі машиналық оқыту алгоритмдерін қолдана отырып, қызметкерлердің өсуін болжау моделдерін құру мақсатында зерттеу жүргізілді. Бұл зерттеуде градиентті бустинг классификаторы, кездейсоқ орман классификаторы және Keras нейрондық желісі арқылы модельденген деректер жиынтығы бойынша эксперименттер орындалды. Күрделі бағалау процесі арқылы қызметкерлердің көтерілуін болжауға арналған машиналық оқыту алгоритмдерінің тиімділігі метрикалық бағалау алгоритмдері арқылы талданды. Зерттеуде алгоритмдерді оқыту мақсатында қолданылған мәліметтер базасы ретінде модельденген қызметкерлер деректері пайдаланылады. Зерттеу барысында модельдердің болжам нәтижелерін көрсету үшін веб-қосымша әзірленді.

Түйін сөздер: машиналық оқыту, рейтинг, нейрондық желілер, терең оқыту, талдау, болжау, персоналды басқару, жоғарылату.

Introduction

Decision about employee`s promotion to the next position has always been an issue in any company. It affects both employee`s motivation to work in the organization and its manager, who has to make such difficult decisions. That is why the problem was analyzed and prediction model was built as a solution.

There are studies aimed at developing employee prediction models. In the work «Prediction of Employee Promotion Based on Personal Basic Features and Post Features», authors focus on testing different machine learning algorithms on predicting employee promotion by building personal basic features and post features based on five strategies [1].

In another research «A Data-driven Analysis of Employee Promotion: The Role of the Position of Organization», the focus of study was to put forward ideas on data-driven solution to the promotion issue in human resource management, and focus on the influence from the position of organization [2].

This paper focuses on building and testing employee`s promotion prediction models and developing software as an interface for better user experience. Employee promotion prediction is a group of models that takes employee`s features as an input and returns possibility of promotion of the employee as an output.

1. Data source

For this research, simulated dataset was created using real HR dataset of employees collected from 2020 to 2021. During this period, approximately 20% of employees were promoted to the next position. The real dataset has 14000 entries and 24 features. The simulated data consists of artificially created 1740 entries. It was further cleaned and the following data processing practices were done:

1) unnecessary features were removed from dataset;
2) temporary employees such as foreign interns were not included to the final set.

After processing dataset, the data has 1500 employees with 8 features (See Figure 1).

	n_of_individual_projects	motivation_r	relationship_with_others_r	communication_skills_r	task_management_r	salary(thousand kzt)	total_rating	promoted
0	2.0	0.200000	0.200000	0.200000	0.200000	100.0	0.800000	0.0
1	3.0	0.500000	0.500000	0.500000	0.500000	150.0	0.700000	1.0
2	3.0	0.907390	0.880383	0.886172	0.854583	86.0	0.717144	0.0
3	0.0	0.954438	0.951636	0.656876	0.959900	441.0	0.947162	0.0
4	1.0	0.677607	0.975069	0.532699	0.451263	385.0	0.458558	0.0
...	...	...	...	...	...	...	...	...
1497	3.0	0.552977	0.943717	0.650876	0.744704	285.0	0.761386	0.0
1498	4.0	0.892194	0.918649	0.918311	0.877594	353.0	0.995365	1.0
1499	3.0	0.886726	0.878922	0.968930	0.981146	372.0	0.854296	1.0
1500	2.0	0.981859	0.677604	0.762169	0.887664	147.0	0.903720	0.0
1501	0.0	0.843148	0.671036	0.788166	0.993719	429.0	0.977252	0.0

1502 rows × 8 columns

Figure 1. Simulated dataset of employees

2. Methodology

In this section, abilities of classical supervised machine learning algorithms and neural network are described.

2.1 Gradient Boosting (GB) is a machine learning technique for regression and classification that is an ensemble of decision trees [3]. This algorithm is based on iterative training of decision trees for minimizing the loss function. Due to the peculiarities of decision trees, gradient boosting is able to work with categorical features and cope with nonlinearities. It uses «boosting» method to convert poorly trained models to well trained ones. Thus, in boosting, each new tree is trained on a modified version of the original dataset.

The idea of gradient boosting originated in the observation by Leo Breiman that boosting can be interpreted as an optimization algorithm on a suitable cost function [4].

2.2 Random Forest (RF) is a set of decision trees. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned [5, 6].

2.3 Keras is a deep learning API written in Python, running on top of the machine learning platform Tensor Flow [7]. Keras was built with a focus on ensuring fast experimentation. Going from idea to result as quickly as possible is essential to doing good research [8].

3. System architecture

System architecture is illustrated in Figure 2. Before applying machine learning algorithms on dataset, data processing was done to the training data. After which the models were developed by analyzing the results of algorithms` work. The concluding step was to develop software (web-application) for users to be able to interact with the built models.

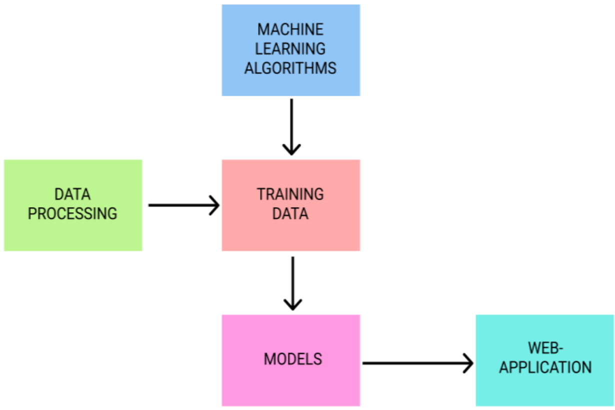

Figure 2. System architecture

Web-application consists of Flask as server Api and React JS as an interface application. Web-application workflow is shown in Figure 3.

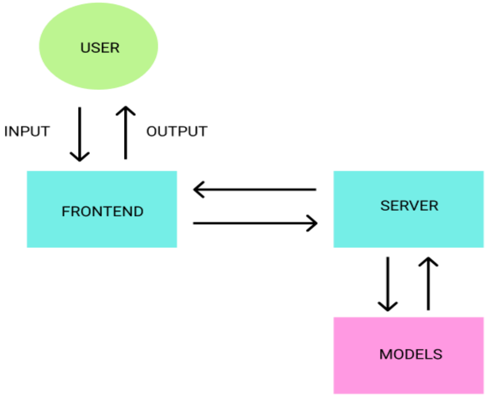

Figure 3. Web-application scheme

The user fills up form with information about specific employee who the decision about its promotion is going to be made. After that, clicking one of the buttons representing different models will output calculated prediction as a result.

4. Results

The results of the mentioned algorithms work on the training dataset, as well as on new samples are presented in this section. In this research, the positive class is assigned to the employees who are promoted, whereas the negative class is for employees who are not promoted. Evaluation metrics and libraries used for this research are as follows:

1) Receiver operating characteristic curve (ROC) and area under the curve (AUC) [9];
2) Confusion Matrix;
3) Classification Report.

ROC and AUC metrics results are displayed in Figure 4. From the results, one can see that all three algorithms showed quite similar values. However, Keras Sequential had the highest AUC value (0.986).

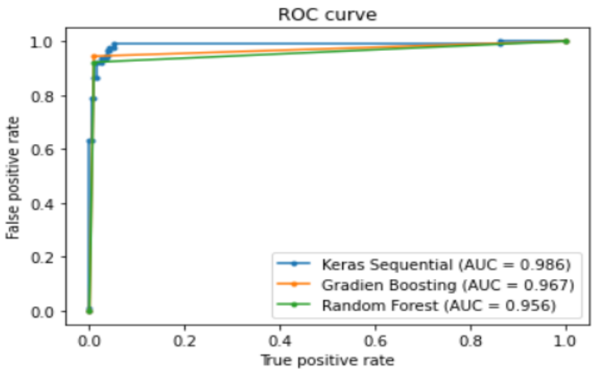

Figure 4. ROC and AUC curve metrics results

The general idea of Confusion Matrix evaluation is to count the number of times instances of class A are classified as class B [10]. Results of confusion matrices (Figure-5, Figure-6, Figure-7) show that all algorithms perform at approximately similar level. Nevertheless, one can notice that Gradient Boosting made less mistakes on predicting promoted employees (only 1 case). Results of classification reports in Figure-8, Figure-9, Figure-10, demonstrate that Gradient Bosting and Random Forest algorithms are evaluated similarly by classification reports. Keras Sequential was assessed with slightly less points on precision and f1-score.

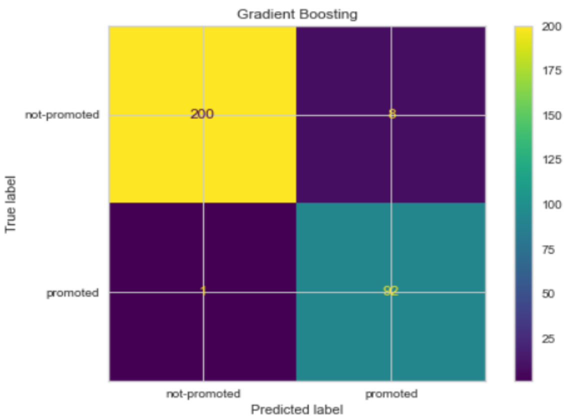

Figure 5. Confusion matrix of Gradient Boosting model

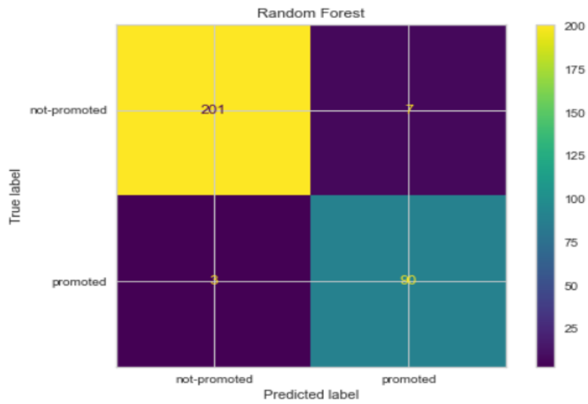

Figure 6. Confusion matrix of Random Forest model

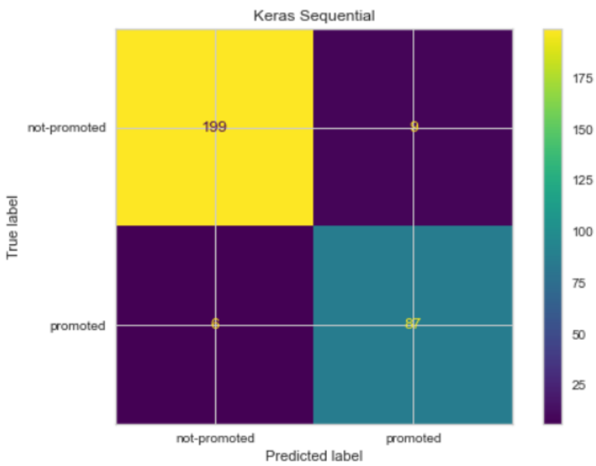

Figure 7. Confusion matrix Keras Sequential model

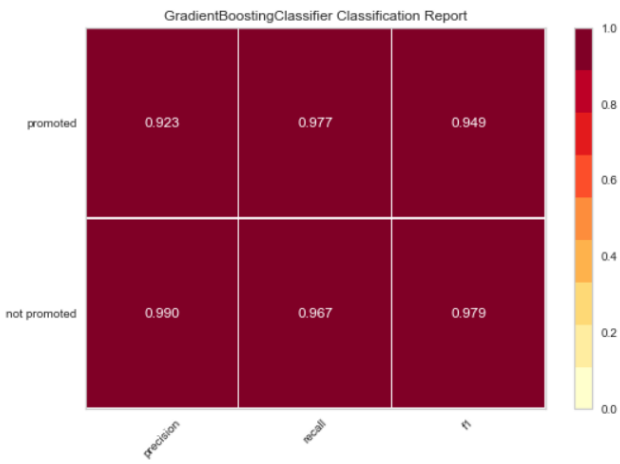

Figure 8. Classification report of Gradient Boosting model

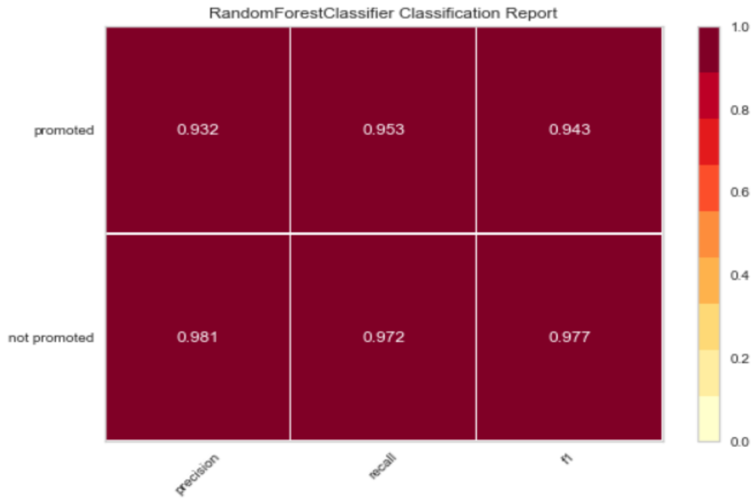

Figure 9. Classification report of Random Forest model

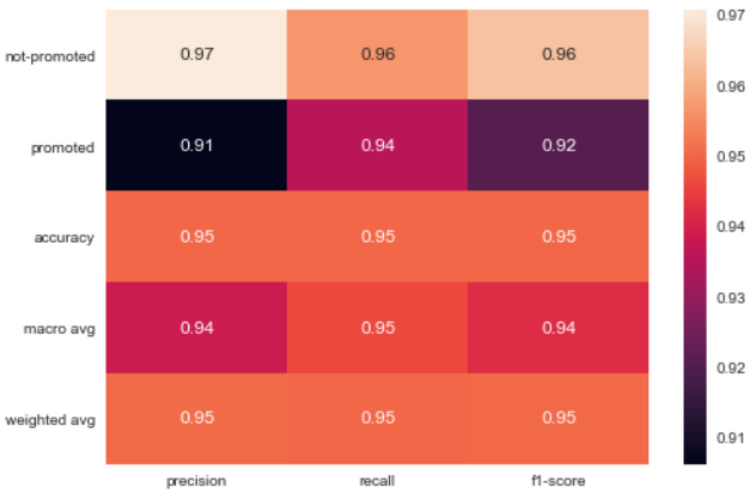

Figure 10. Classification report of Keras Sequential model

Prediction on new samples were done using developed web-application. Figure 11 illustrates a sample employee`s information and promotion predictions of different models.

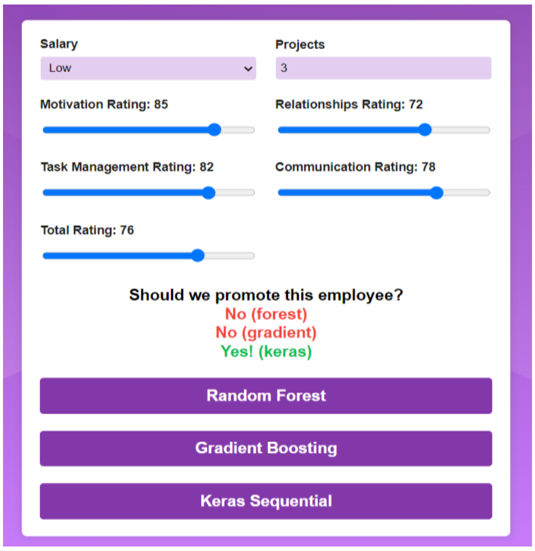

Figure 11. Results of models on new input

For the new inputs, GB and RF algorithms perform stably and similarly. When it comes to Keras model, it can show anomaly behaviors such as if one of the features are too low or too high, it can still predict positive (promoted).

Conclusion

During this research work, it has been identified that prediction of employee promotion is very important aspect in human capital building process. Three popular machine learning techniques for forecasting employee`s promotion were built and analyzed.

Training dataset of employees were created using real data of companies and further processed. Popular supervised machine learning algorithms were studied and tested on performance using numerical metrics. The metrics, used in the research, showed that all three models were able to predict with close accuracy, despite small differences.

To test the models for new samples, special software was developed, by using which it is possible to predict the advancement of employees. These tests showed that the Gradient Boosting and Random Forest models gave the same predictions, while the Keras model gave different answers in most cases, which suggests that the Gradient Boosting and Random Forest models are preferred over the Keras model for predicting employee progress.

Acknowledgments

This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant #AP09259208).

References:

1 Long, Yuxi & Liu, Jm & Fang, Ming & Wang, Tao & Jiang, Wei. "Prediction of Employee Promotion Based on Personal Basic Features and Post Features". Proceedings of the International Conference on Data Processing and Applications (2018): 5-10. https://doi.org/10.1145/3224207.3224210

2 J. Liu, T. Wang, J. Li, J. Huang, F. Yao and R. He. "A Data-driven Analysis of Employee Promotion: The Role of the Position of Organization". IEEE International Conference on Systems, Man and Cybernetics (SMC) (2019): 4056-4062. https://doi.org/10.1109/SMC.2019.8914449

3 Piryonesi, S. Madeh; El-Diraby, Tamer E. (2020) "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of Infrastructure Systems: 04019036. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000512.

4 Breiman, L. "Technical Report 486". Statistics Department, University of California, Berkeley (1997): 11-13

5 Ho, Tin Kam. "Random decision forests". Proceedings of the 3rd International Conference on Document Analysis and Recognition (1995): 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

6 Ho TK. "The Random Subspace Method for Constructing Decision Forests". IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (8) (1998): 832–844. https://doi.org/10.1109/34.709601.

7 "About Keras", Official website of Keras Sequential Api, Google Keras team, accessed December 1, 2021, https://keras.io/about/.

8 Deng, H.; Runger, G.; Tuv, E.. "Bias of importance measures for multi-valued attributes and solutions". Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN) (2011): 293–300. https://doi.org/10.1007/978-3-642-21738-8_38.

9 Altmann A, Toloşi L, Sander O, Lengauer T. "Permutation importance: a corrected feature importance measure". Bioinformatics. 26 (10) (2010): 1340-7. https://doi.org/10.1093/bioinformatics/btq134. PMID 20385727.

10 Géron, A. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. (O'Reilly Media, 2017). 92-93