

DOI: [10.32702/2307-2105-2021.1.152](https://doi.org/10.32702/2307-2105-2021.1.152)

УДК 004.8:005.95

*D. Yurchenko,
Master's degree, Borys Grinchenko Kyiv University, Kyiv
Data Scientist, RBC Group, Kyiv
ORCID ID: 0000-0002-3900-0476*

USING MACHINE LEARNING TO OVERCOME THE PETER PRINCIPLE IN MODERN HIERARCHICAL ORGANIZATIONS

*Д. В. Юрченко,
магістр, Київський університет імені Бориса Грінченка, Київ
Data Scientist, "АрБіСі Груп", Київ*

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ПОДОЛАННЯ ПРИНЦИПУ ПІТЕРА У СУЧАСНИХ ІЄРАРХІЧНИХ ОРГАНІЗАЦІЯХ

According to the Peter principle, in hierarchical organizations all workers sooner or later become incompetent due to being promoted to positions that don't suit them. That is caused by traditional promotion strategies that favor the most successful employees while it does not always happen to be the best approach.

On the other hand, applying unconventional promotion policies might result in negative psychological effects, including drop in personnel motivation. General promotion rules also ignore the possibility that the Peter principle may hold for some hierarchical pathways while remaining inactive for other types of promotions.

The purpose of this article is to justify the new way of preventing the Peter principle effect by employing machine learning (ML) models to predict workers' potential competence level they would acquire in case of promotion.

The article proposes the algorithm of organizational promotions based on the ML approach. The set of employees' characteristics for the models to be trained on, as well as specific ML algorithms, namely linear-based and tree-based are suggested. Inner working specifics of the algorithms are detailed from the mathematical point of view. The principles of the modern ML training process are described and the training data requirements are highlighted.

The advantages of employing a dynamic personalized promotion strategy based on machine learning methods are described. Such a strategy allows treating separate promotions individually, choosing from candidates solely on the basis of the model prediction, thus avoiding the unnecessary usage of unconventional policies and subsequent motivational drops. In addition, ML-based approach opens up possibilities to substantiate promotion choices, as those were derived from actual historical data used for training the model. This excludes a potential sense of prejudice while using multiple features for prediction eliminates the risk of unscrupulous employees' behavior aimed at manipulating their own performance with the purpose of receiving a promotion.

Properly implemented, the proposed algorithm can help to eliminate the negative effect of the Peter principle while mostly preserving the motivational component of promotions.

Відповідно до принципу Пітера, у ієрархічних організаціях всі працівники рано чи пізно досягають свого рівня некомпетентності внаслідок підвищення на посади, які їм не підходять. Таке явище спричиняється традиційними стратегіями просування по службі, котрі передбачають вибір для підвищення найбільш успішних працівників, хоча такий підхід не завжди виявляється найкращим.

З іншого боку, застосування нетрадиційних політик просування може призвести до негативних психологічних наслідків, включаючи падіння рівня мотивації персоналу. Узагальнені правила підвищень також ігнорують можливість того, що принцип Пітера може діяти при підвищеннях між окремими ієрархічними ланками, залишаючись при цьому неактуальним для інших типів підвищень.

Метою статті є обґрунтування нового способу запобігання принципу Пітера через використання моделей машинного навчання (МН) для прогнозування потенційного рівня компетентності працівників, якого вони набудуть у разі підвищення.

У статті запропоновано алгоритм організаційного службового просування на основі використання машинного навчання. Запропоновано набір характеристик працівників, які можуть бути використані для навчання моделей, а також конкретні алгоритми МН – лінійні та засновані на деревах. Специфіка роботи алгоритмів деталізована з математичної точки зору. Описано принципи навчання сучасних моделей та висвітлено вимоги до навчальних даних.

Описано переваги використання динамічної персоналізованої стратегії підвищень на основі методів машинного навчання. Така стратегія дозволяє розглядати окремі підвищення індивідуально, обираючи серед кандидатів виключно на основі прогнозу моделі, і тим самим уникати необґрунтованого використання нетрадиційних політик та пов'язаних з ними мотиваційних втрат. Крім того, підхід, заснований на МН, відкриває можливості для обґрунтування вибору працівника на підвищення, оскільки цей вибір був продиктований фактичними історичними даними, використаними для навчання моделі. Це виключає виникнення у працівників почуття упередженості, а використання кількох факторних ознак для прогнозування усуває ризик недобросовісної поведінки працівників, спрямованої на маніпулювання результатами власної праці з метою отримання підвищення. За умов його належної реалізації, запропонований алгоритм може допомогти усунути негативний ефект принципу Пітера, при цьому більшою мірою зберігши мотиваційний компонент підвищень.

Keywords: *Peter principle; promotion strategy; machine learning; posterior competence; hierarchical organization.*

Ключові слова: *принцип Пітера; стратегія підвищення персоналу; машинне навчання; апостеріорна компетентність; ієрархічна організація.*

Stating of the problem in general terms and definitions of its interrelation with important scientific or practical tasks. First introduced in 1969 by Dr. Lawrence Peter and Raymond Hull [1], the Peter principle is known as a managerial hypothesis referring to the phenomenon of gradual hierarchical degradation caused by traditional personnel promotion strategies which are based on the rule of favoring the most successful employees. The Peter principle suggests that while companies tend to promote those who showed the best results at their current positions, this might not be the optimal decision as higher positions often require a vastly different set of hard and soft skills that employees being promoted might not possess.

This leads to the poor performance of newly promoted employees as their competence simply does not transmit to the next positions. For various reasons employees who have reached their level of incompetence are not demoted (and neither are promoted further) afterward, but instead get stuck, not being able to meet the requirements of their current positions and thus ending up on “Peter’s Plateau” where they face “Final Placement” syndrome [1, p. 24]. Since promotion rules are usually universal across all the organization, the Peter principle eventually finds its effect in the drop in organizational effectiveness and thus, might result in substantial financial losses [1]. Potential disastrous consequences of such a phenomenon force to search for alternative approaches to organizational management.

Analysis of recent investigations and publications which initiated the solution of the stated problem and on which the author relies; highlighting unsolved issues of the overall problem the article is dedicated to.

Numerous studies were carried out since the seventieths in the attempt to confirm or refute the Peter principle in real-world organizations, with probably the most significant one presented in 2019. The extensive research titled “Promotions and the Peter Principle” has been conducted by the professors of MIT A. Benson, D. Li, and K. Shue [2]. The analysis was based on actual employee performance data provided by 131 US sales companies from the field of IT, manufacturing, professional services, etc. The total amount of workers covered was almost 39 thousand; the analysis time period was from 2005 to 2011.

The paper shows that firms consistently promote top salespeople to managerial positions even though these promoted workers eventually make worse managers than those who had managerial potential but performed worse in sales [2, p. 39]. The authors further note that such policy leads to the fact that subordinates of the managers unsuitable for their roles perform 30% worse compared to their performance under the policy of promotion which ensures the best role matching. In particular, negative relationship between sales and management quality of the same employee can be expressed in the following terms: doubling the sales of the employee corresponds to the drop in their value added as a manager by 6.1% [2, p. 17]. This is direct evidence of the Peter principle effect.

As for ways of dealing with the consequences of the Peter principle it was suggested to apply different promotion policies, those that do not favor the most successful employees. One of the most notable studies was published by Italian scientists A. Pluchino, A. Rapisarda, and C. Garofalo of the University of Catania in 2010 (it was awarded the Ig Nobel Prize in management) [3].

It adopted agent-based modeling to simulate organizational efficiency with respect to the type of promotion strategy (promoting the best employees, promoting the worst employees, promoting randomly), concluded that given that Peter’s principle holds it is the most rational to promote the worst and suggested to promote employees randomly in general, supposing one doesn’t know whether the hypothesis holds beforehand [3, p. 6-9].

Despite receiving a lot of critique for not representing organization as a social, dynamically intelligent structure and thus invalidating the results, the Italian paper served as an inspiration for several similar studies [4; 5; 6; 7], none of which however managed to produce conclusive solutions for overcoming the Peter principle.

The aims of the article (stating the objective). The purpose of the article is to propose a new way of dealing with the Peter principle in modern hierarchical organizations through the usage of machine learning algorithms for predicting employees’ performance at higher positions and thus eliminating the cases of promotions that lead to reaching employees their level of incompetence.

The wording of the main material of the research with comprehensive argumentation of obtained outcomes. One of the key issues regarding using alternative promotion strategies as an attempt to overcome the Peter principle was their negative psychological effect, including the loss of the motivational component of promotions. In organizations promotions are used to serve two roles – firstly, providing the best employee-position matching, and secondly, ensuring personnel motivation through career opportunities [8, p. 364]. Should it be chosen to promote anyone but the best, such decision will at least partially eliminate motivation stimulus not only for the employee who failed to receive promotion but also for the rest of the personnel as other employees are able to detect and follow reinforcement rules, seeking for behavior that guarantees the best reward (namely, maximizes promotion probability). Specifically, employees will probably be aiming to reach the same level of performance, no matter high or low, that has previously ensured promotion.

Needless to say, it can’t be allowed to lower overall motivational level, that’s why radical propositions as the one choosing the worst performing workers for promotion are considered naïve and inapplicable in actual social environments [5, p. 59]. Moreover, firms seem to be aware of the Peter principle phenomenon but follow traditional promotion policies nevertheless, seeking to exploit career-based incentives even at the cost of potential lowering managerial quality [2, p. 39].

One of the possible solutions presented in this regard was to apply a “double standard” promotion policy [6, p. 11]. In plain terms, it’s just a method of continuous disinformation that misleads personnel into thinking that best employees are being promoted while in fact, it’s the worst who are receiving higher offices. Such condition, should it be properly manufactured, is expected to legitimize conclusions made by Pluchino et al. by the means of reducing complex adaptive social nature of real hierarchical organization to the non-learning, never-changing deterministic system of agents.

However, we believe such an approach is unrealistic in a long-term perspective as it seems impracticable and unhealthy to maintain a constant state of disguise, especially in large organizations where the risk of information leakage is great and potential consequences of such disclosure would undoubtedly have a ruinous effect of the present and future of company’s state of affairs.

Following the direction of thinking previously established in scientific literature, we propose forming a personalized dynamic promotion strategy that consists of choosing candidates for promotion based on the predictions retrieved from a trained machine learning model.

Machine learning (ML) is a subfield of artificial intelligence (AI) that relies on automatic statistical data processing which allows algorithms to capture patterns and dependencies present there, learning from them and using gained knowledge to extrapolate to new data later, making predictions (this is called inference).

For the last 10 years, ML has been vastly applied both in forms of classical statistical algorithms (e.g. linear regression) and deep learning (DL) which is represented by multi-parameter multi-layer artificial neural networks (ANN) able to model complex relations and achieve groundbreaking results in many areas of science, technology, and business.

Nowadays machine learning and data science are an essential part of financial, operational, and marketing analysis in various types of organizations worldwide and in Ukraine, and an integral component of most modern online and app services [9]. They also help in developing solutions in neuroscience, biology, healthcare, etc. For example, a recent breakthrough in predicting the structure of proteins, a problem that had remained unsolved for 50 years, is a merit of a deep neural net called AlphaFold 2 [10].

Arguably the largest part of today's ML research falls into the category of so-called supervised learning – a type of learning that requires providing a dataset that is structured in a features-labels manner (the other two categories are unsupervised and reinforcement learning). After being trained to predict given labels from given features, the ML algorithm is fitted on previously unseen data (features) to predict labels for it.

In such fashion, predictions made by ML models could be used to estimate posterior employee competence, i.e. the competence that employee would have acquired in case of promotion. Naturally, applying ML requires dealing with two issues – choosing algorithm type and data to train it on. Let's elaborate on each of those issues below, starting with the latter. As was mentioned, our goal is set to predict employee performance level on a higher hierarchical position so that we could make an informed decision about filling the position. In order to properly do that it is first needed to estimate which known and measurable factors could potentially have an influence on the target value.

In its core, any ML algorithm aims to approximate a function that transforms these factor variables into predicted outcome by analyzing provided (historical, i.e. already known, or at least ready to be provided as needed) features-labels correspondences – training data points. It does so by optimizing an objective function that typically encompasses the difference between model outputs and true labels. In our case, the label variables values could be real numbers on a scale from 0 to 10 representing employee posterior competence while any employee characteristics that could (partially) predetermine their future performance should serve as features. We propose for consideration the next set of features:

1. Current employee level of competence (prior to potential promotion) measured on the same scale as target (posterior to promotion) level of competence
2. Current employee hierarchical level
3. Potential employee hierarchical level
4. Current employee job type (full-time/part-time/contract)
5. Potential employee job type
6. Current employee hierarchical level type (linear/managerial)
7. Potential employee hierarchical level type
8. Current employee commuting type (in-office/remote)
9. Potential employee commuting type
10. Current employee salary
11. Potential employee salary
12. Specialization/Job title
13. Age
14. Experience level
15. Education level
16. Gender
17. Marital status (single/married/divorced/widowed)
18. Familial status (yes/no)

It is vital to understand that the proposed list is in no way universal for all organizations. It should be adjusted (completed or reduced) by human resources workers or managers based on the specifics of each organization and the circumstances that define an employee holding a given position at a given time. For example, it might be justified to include the time of the promotion if promotions are not conducted at a fixed frequency in the organization, or exclude marital/familial status should this characteristic be recognized as an invasion of privacy, etc.

That way, the algorithm proposed for choosing optimal candidates for promotion through the usage of machine learning models is as follows:

1. Forming dataset containing prior and posterior competencies with unique characteristics of every promoted employee and properties of promotion while conducting regular promotion strategy normally established in organization. This is considered as a preparation period during which promotion rules remain unchanged, but measurements are being taken in order to acquire data needed to train a ML algorithm on.
2. Training an ML regression algorithm on the gathered data of historical promotions.
3. Using a trained ML model for inference, obtaining prediction of the posterior competence value of the candidates for promotion and making a managerial decision on that basis – e.g. what candidate to choose from company's own employees, or whether to invite a specialist from outside of the organization, etc.
4. Continuing gathering data and retraining the model after each promotion to take into account updated information.

In general, the more data is accessible to the algorithm (and the more diverse it is), the lower the model test error would be (meaning the model would be less prone to overfitting to training data); hence it is essential to never stop feeding new data, allowing the model to capture current regularities in employees competence transmission during promotion. These regularities might depend on the specific places in organization where promotions take place. For example, competences might vary more significantly for promotions from the 4th to the 3rd hierarchical level but remain

approximately the same for the 3rd – 2nd pair of levels because of the greater structural differences between the former pair: say, the 4th is a level of linear workers and the 3rd is the lower management level while the 3rd and the 2nd are both management-related levels which makes them closer to each other in terms of the skillset required by positions they contain. This is actually an extremely important part of the dataset forming as the effect of the Peter principle is probably the most noticeable during promotions to managerial positions from non-managerial ones (see research by Benson et al.).

It is understood that while ML algorithms learn best when provided with large and diverse datasets, the proposed promotion strategy does not imply using already existing organizational data (unless described measurements have actually taken place in the past), thus making it problematic to prepare the data needed – the majority of the organizations willing to improve their promotion strategies are simply not large enough and don't have personnel career growth intensive enough to come up with such a dataset in a reasonable period of time.

Even an approximate number of promotions required to form a trustable dataset is hard to determine unconditionally (although it is believed to have to be significantly larger than the number of features used in the model for avoiding the so-called “curse of dimensionality”), but assuming that promotions may take place once every year or even more rarely, it is probable that small businesses would consider the very first step of the proposed algorithm somewhat overwhelming.

As a possible solution to the data limitation issue, it could be proposed to derive feature values from the historical records. If the existing promotion strategy involves some tangible data being gathered for promotion basis (i.e. not only organoleptic experience of direct managers matters) – say, sales records for sales managers or number of patients served for medical personnel – the history of such measurements could be used as a prior competence level estimate granted it is converted to the same scale as a posterior competence for compatibility. The posterior competence level in turn would be measured as-is for each of the previously promoted workers still occupying the position they were promoted to. The values of the rest of the features at the time of promotion (age, experience period, education, salary, etc.) could be easily restored from historical data.

The second part of the ML-based solution refers to a choice of the algorithm type. In this context, it should be stated that current machine learning is a field that is inherently peculiar to a great deal of uncertainty and ambiguity. This is especially problematic for large neural networks which are often referred to as black boxes for their lack of interpretability. It is often hard to estimate which of the ML algorithms would suit a given problem the best (especially if the problem has not been addressed before), and since nowadays these are prevalently available in the form of high-level API organized in programming packages (modules), it is more time-efficient to just employ models one by one and see which one produces the lowest test error. That is, the issue of choosing a specific algorithm is considered to be rather empirical than a theoretical one.

For convenience though, we propose two algorithms that seem to be reasonable starting points for a given problem.

Linear regression is arguably the simplest ML algorithm that had been used as a statistical and econometrical method long before machine learning saw its rise as a separate discipline. Its basic premise is that the dependent variable (target value, i.e. label) linearly depends on the independent variables (features). Thus, a linear regression takes the form:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon, \text{ where} \quad (1)$$

y – target variable (posterior competence level being predicted);

$\beta_1, \beta_2, \beta_n$, and ε – learnable model parameters: intercept terms and the error term;

x_1, x_2, x_n – feature variables (employee and positions characteristics).

In ML the learning process of linear regression is as follows:

1. Initialization of model parameters as zeros or small random values and calculation of the function value representing competence of the employee at the next level position.
2. Computing the error between the calculated function value and the actual value of a posterior competence (usually the mean squared error is used). This is needed to interpret model performance, how well it has evolved as the result of training.
3. Iterative adjustment of parameters via backpropagation of the error through the gradient descent, i.e. by changing the parameters by the value of the partial derivatives of the error function with respect to each of the model parameters with the purpose of minimizing the error value.

All calculations listed above are made automatically in a matter of seconds through the usage of ML packages for Python or R programming languages (e.g. scikit-learn is one of the most common). After training is completed, the model should be able to approximate the function of the relationship between the provided features and the posterior competence of the employee, and then it will be enough to simply substitute the values of feature variables (characteristics of candidates) before the next promotions to obtain the value of competence after potential promotion, and in that way select the best candidate.

For instance, consider two candidates for promotion conducted from the 3rd level to the 2nd level of a medical institution (Table 1). Then once the model is fitted on historical data all we need to do is apply it to all the listed features

in order to get predicted values of competence for both candidates. We can then choose the one with the highest predicted competence.

Table 1

Proposed set of features for the ML model training (generated by the author)

Features	Candidate 1	Candidate 2
Current level of competence (scale of 10)	8.2	7.9
Current hierarchical level	3 rd	3 rd
Potential hierarchical level	2 nd	2 nd
Current job type	Full-time	Full-time
Potential job type	Full-time	Full-time
Current hierarchical level type	Managerial	Managerial
Potential hierarchical level type	Managerial	Managerial
Current commuting type	In-office	Remote
Potential commuting type	In-office	In-office
Current salary	\$65 000	\$70 000
Potential salary	\$80 000	\$80 000
Specialization/Job title	Head of the Department of Transplantology	Head of the Consultative-diagnostic Department
Age	37	35
Experience years	15	11
Education level	Ph.D.	MD
Gender	Female	Male
Marital status	Divorced	Married
Familial status	Yes	No

Note how the proposed ML-based method allows to automatically take into account the potential effect of the Peter principle for each individual promotion instance rather than imply that it either holds or doesn't hold for the whole organization and all the employees in general. This points to the key idea of this paper: the need for the usage of statistical algorithms to come up with a decision in each separate case by taking into account its unique attributes that have an inherent influence on the resulting variable.

Besides the linear relationship between variables, linear models make certain assumptions about relations present in the data they use for training and inference: the absence of correlations in features values, homoscedasticity, and independence of errors – however, in practice it is almost never possible for these assumptions to hold true. For example, both the experience and age seem to be important regressors (the former denotes professional readiness while the latter possesses information about general current and future psychological behavior at work) it is obvious they must have a strong positive correlation (same goes for experience and salary level, etc.).

In order to be used as model training data, all feature variables values should be converted to numeric form. One minor issue which, however, can have a great negative influence on the prediction results is the exact encoding fashion, meaning the way to convert feature values from text to numeric. The main challenge here is a loss-free transfer of the logical structure of features into numeric form.

For example, if we choose to encode different levels of education, we should ensure that greater values correspond to higher levels (Ph.D. should be treated as superior to MD), but at the same time, we shouldn't trick the algorithm into thinking that the same principle applies for categorical features like marital or familial statuses. There is naturally no direct order of superiority for single/married/divorced/widowed statuses, but for simple numeric encoding, a divorced candidate would be treated better by the model than a single one the same way as males would be preferred over females or some job titles over the other ones from the same hierarchical level.

To eliminate that bias different encoding techniques were proposed, one of which is one-hot encoding (OHE). It implies splitting a feature into several binary features, of which only one would be denoted as 1 while values of the rest would be set to 0. The number of features after such transformation represents the cardinality of the original feature, meaning the number of unique values this feature could acquire. So, given the marital status variable can have 4 possible values, one-hot encoded marital status variable would be split into 4 separate features (columns), with 3 of them equal to zero, and one equal to 1. This way, employees with different statuses won't be numerically comparable to each other but will rather be treated as just differing. However, this method obviously has disadvantages in terms of dimensionality expansion and is not suitable for a large number of categorical features of high cardinality.

There are various ways to encode the data, as well as to automatically choose the most significant features (so-called dimensionality reduction – see correlation matrices, principle component analysis, t-distributed stochastic neighbor embedding, and other methods), but they are out of the scope of this article.

The other approach we propose in terms of algorithm selection is tree-based rather than linear-based. Namely, one could make use of random forest or XGBoost algorithms. Random forest is a bootstrap aggregating (so-called bagging) ML algorithm while XGBoost is an implementation of gradient boosting technique, however, both of these are known as ensemble modeling instances that make use of decision trees as their internal building blocks. Now, decision trees are simple non-parametric predictive models that utilize the principle of the best split in order to hierarchically

solve a regression or classification task. Basically, the idea behind those is top-down recursive partitioning of the data that results in the formation of tree pathways comprising of the values of different features with all possible target variable values (intervals) at the very bottom.

The final result of the decision tree algorithm is obtained by taking into account each feature variable and splitting the data on this basis according to the “true-false” test. For example, the model has started off with evaluating the “age” attribute of a candidate for promotion as a feature that provides the best (the cleanest) separation, and then it iteratively climbs down along this branch, splitting the target data space (a posterior competence values) depending on the other feature variables. By the best split, it is usually meant the one that provides the greatest reduction in the value of Gini Impurity.

The final prediction value is the average of all values of a posterior competence for previously promoted employees whose attributes values belonged within the values in the nodes of the tree (higher or lower than a certain age, higher or lower than a certain value of a prior competence, etc.). A graphical example of a simple regression decision tree is shown in Figure 1. It demonstrates values being predicted, a number of training samples, features chosen for split ($X[2]$, $X[4]$), and the split error value (MSE, mean squared error).

Random forest simply comprises a certain amount of decision trees (ranging arbitrarily from several to several thousand) while its prediction is usually arithmetic mean or mode of all its trees; also, only a random sample of features and training points is used. This approach allows preventing overfitting to the training data by leveraging the decision among several voters and different areas of data.

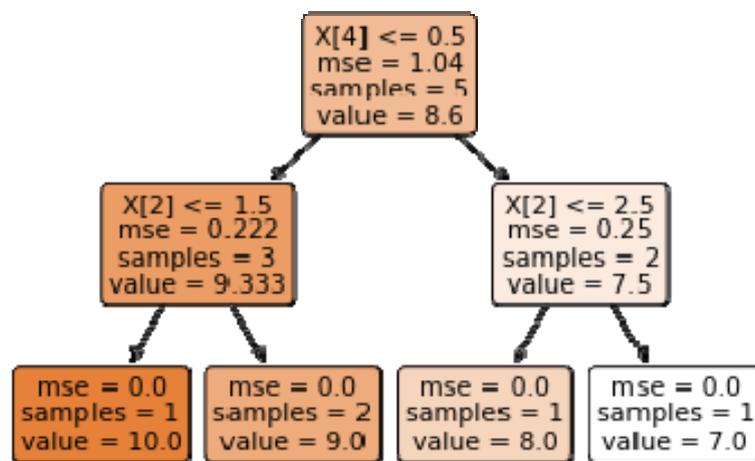


Figure 1. A regression decision tree structure
(generated by the author)

Gradient Tree Boosting (GTB), probably the most popular implementation of which is XGBoost model, introduced in 2014, is a somewhat similar, but more complex and sophisticated algorithm. GTB helps to make each next decision tree added to the model fix the mistakes (lower the error) left by the previous tree. It is done by sequentially fitting next trees on the derivative of the loss function with respect to the outcome of the previous ones, so-called base or weak learners, correcting previous outputs with gradients for each node of the tree. Generally, the same idea which is used for training linear regression or neural networks applies here, except instead of adjusting separate existing parameters of the model, GTB relies on outputs of decision trees, and learning is conducted by adding new models to the ensemble.

Currently, XGBoost is recognized by many field researchers as a gold standard in the world of classical ML algorithms and is advised to be used “when in doubt” [11].

Everything mentioned above that refers to training, practical implementation, and features needed for linear regression applies to both of the described tree-based algorithms as well. Although modern deep neural networks are extremely powerful instruments that have reached state-of-the-art results in natural language processing and generation, computer vision, time-series forecasting, recommender systems, playing board and computer games, and many other sets of challenges, and currently are believed to be the head of the curve in AI, we omit considering these methods for a given problem. The main reason for that is data requirements – ANNs are generally greedy for the dataset size which as we have already stated might become a time-related issue for some companies.

Now, compared to the general change in promotion rules, the machine learning approach provides more flexibility and opens opportunities to soften motivational losses mentioned at the beginning of the article. As was mentioned, contrary to changing strategy all over the organization, the proposed approach implies predicting competence level for each individual promotion. Firstly, that allows not to spoil the cases where it’s still reasonable to promote the best workers and such wise to avoid unnecessary motivation losses.

Secondly, the absence of a general promotion rule leads to workers not being aware of their promotion chances beforehand since now they can’t judge (and adopt their behavior) only based on their current performance – it’s only in

tandem with other attributes that their posterior competence could be predicted and promotion decision could be made. And it's improbable to guess which values these attributes should acquire in order to ensure promotion, which is basically the idea behind ML – to capture patterns that humans can't. Moreover, the prediction is derived solely through a programmatically automated mathematical algorithm, which is a pre-trained ML model. Because of all of that, it seems unlikely for employee performance or other characteristics to become an object of unscrupulous manipulation with the purpose of receiving a promotion.

Crucially, ML brings transparency to the table: now managers can explain to their successful subordinates why they have been rejected promotion (assume Peter hypothesis holds). This requires certain communication skills but substantiating that given all the circumstances an employee would perform worse at a higher position and it would reflect their own professional well-being might be more effective than bluntly pointing out the abstract idea that for some reason organization currently needs the worst employees to be promoted while the best performers are expected to demonstrate trust and loyalty.

If one manages to intelligibly deliver a message, convincing employee that it's in their own best interest to abandon promotion, stressing that it's highly likely that employee competence would be lowered once they're promoted (and that it's probably no one's fault), it seems probable to retain the same or reasonably lower level of motivation. Operating actual data and relying on statistical evidence would look more convincing and at least would eliminate the risk of employee getting the impression of personal prejudice as all calculations are being made strictly through standardized computer code processing historical data from their very organization.

However, we still tend to agree with American researcher Eric Chan who suggests that there's no effective promotion policy that would ensure best position matching while fully preserving motivation induced by career opportunities [7, p. 76]. Opposing vectors of career promotion substance just seem to be interlinked too deeply to safely exclude one of them and retain a positive effect of another.

There is, however, a possibility that given the availability of the large amount of diversified data related to the organization staff, as well as the long history of promotions, ML algorithms could be adjusted to the problem of binary classification that doesn't rely on exploiting prior and posterior competence assessment. That is, by gathering enough features for each of the many previously promoted employees and labeling the promotions as being successful or unsuccessful (according to some preset threshold), it seems technically plausible to predict the success of future promotions only on the basis of open information, without the need for direct evaluation of employee's performance.

There would be no formal division into "good" and "bad" employees then, and the ones who were "good" under the traditional system (that with performance assessment) would be less subject to demotivating processes since they would not possess knowledge about how their exact level of competence compares to those of other candidates for promotion. But at this point, such an option is seen as purely hypothetical as the current competence level is considered to be the strongest predictor for future performance. Additionally, simplifying the problem of 0 to 10 scale regression to binary classification will inevitably lead to a situation where two or more candidates are simultaneously found suitable or unsuitable for promotion making it impossible to distinguish them fairly.

In this context, an attempt should be mentioned to propose a similar ML-based solution exactly for the purpose of overcoming the Peter principle. Contemporary French AI researcher and blockchain specialist Thomas Ferry has recently proposed in his blog to use an ML model to predict an employee's chances of next promotion under the condition they will be promoted now [12]. That is, one's success at the current workplace, according to Ferry, could be measured by whether they are chosen for promotion, hence, by predicting whether the next promotion would take place we predict whether it is plausible to promote an employee at a current timestep. However, algorithms and feature variables, as well as ways to deal with promotions to the final hierarchal level where it happens to be no variable to predict were not specified.

Although we propose a much more detailed and, hopefully, valid solution, it too contains a certain level of arbitrariness with regard to the final algorithm choice. Which is, once again, not the key issue and probably shouldn't be addressed on a theoretical level. One could go with linear regression (or its variants including multinomial regression), random forest, XGBoost, support vector machines, naïve Bayes, KNeighbors classifier/regressor, SGD regressor, and so forth, or even non-ML algorithms like Monte Carlo simulations. The only crucial principle that we urge to follow in this regard concerns making use of the past diversified data that holds patterns defining how well a candidate for promotion will perform.

With modern big data systems (like BI tools) capable of tracking various indicators and stacking large amounts of information, which provides managers of organizations with great analytical opportunity, there's no more need to generalize and oversimplify sets of dynamic varying processes, treating them as one big pile for which intuitive approach is suitable enough. Today, in order to prevent the Peter principle from affecting an organization, managerial decisions must be well informed and data-driven, and machine learning is one good way of ensuring that is the case.

The conclusion is drawn from the research and prospects of further investigation in the respective direction. In order to eliminate the negative effect of the so-called Peter principle in hierarchical organizations, it was proposed by scientists to adopt promotion policies that favor the worst candidates. Such an approach leaves best-performing workers where they are the most helpful while providing more suitable positions for those who perform poorly. However, theoretical positive effect poorly transitions into practice as mentioned strategies risk to be seen by personnel as unfair and lead to the drop of the motivational level.

To overcome the Peter principle and mitigate the mentioned negative effects it was proposed to make use of machine learning models trained on real organizational data to predict employees' performance level after potential promotion.

Properly implemented, described propositions could potentially decrease the probability of promotions which lead to "Final Placement" syndrome while maintaining the same or reasonably lower level of personnel motivation. Firstly, the proposed approach would allow for greater flexibility through addressing only those specific cases where the Peter principle has its negative effect. Secondly, using actual data as statistical evidence not only makes promotion decisions more informed for managers but also justifies it in the eye of candidates for promotion, opening opportunities to soften the negative effects of unpopular decisions to promote poorly performing employees.

Further research on the ways of preserving promotion-based incentives while avoiding the Peter principle is needed. Also, an important step in regard to employing an ML approach would be to empirically test different scenarios of data collecting, features selection, and models training.

References.

1. Peter, L. J. and Hull, R. (1969), *The Peter Principle: why things always go wrong*, William Morrow and Company, New York, US.
2. Benson, A. Li, D. and Shue, K. (2019), "Promotions and the Peter principle", *Quarterly Journal of Economics*, vol. 134 (4), pp. 2085–2134.
3. Pluchino, A. Rapisarda, A. and Garofalo, C. (2010), "The Peter Principle Revisited: A Computational Study", *Physica A*, vol. 389 (3), pp. 467-472.
4. Fetta, A.G. Harper, P.R. Knight, V.A. Vieira, I.T. and Williams, G.E. (2012), "On the Peter Principle: An agent based investigation into the consequential effects of social networks and behavioural factors", *Physics A: Statistical Mechanics and its Applications*, vol. 9 (1), pp. 2898-2910.
5. Fiedor, P. (2015), "The Social Dynamics of the Peter Principle", *Journal of Engineering Science and Technology Review*, vol. 8 (1), pp. 56-60.
6. Olsson, E. J. and Proietti, C. (2015), "Promoting the best as an incentive: reply to Pluchino et al. on the Peter Principle", DOI: <https://doi.org/10.13140/2.1.2201.8888>
7. Chan, E. W. (2015), "Promotion, Relative Performance Information, and the Peter Principle", Ph.D. Thesis, Business Administration, University of Pittsburgh, Pittsburgh, US.
8. Milgrom, P. and Roberts, J. (1992), *Economics, Organization and Management*, Prentice Hall, New York, US
9. SAP (2018), "Robota na avtomati: iak mashynne navchannia pidvyschuie prybutkovist' biznesu", available at: <https://news.sap.com/ukraine/2018/05/machine-learning/> (Accessed 4 January 2021).
10. Fortune (2020), "Lessons from DeepMind's breakthrough in protein-folding A.I.", available at: <https://fortune.com/2020/12/01/lessons-from-deepminds-a-i-breakthrough-eye-on-ai/> (Accessed 4 January 2021).
11. Towards Data Science (2019), "XGBoost Algorithm: Long May She Reign!", available at: <https://towardsdatascience.com/https-medium-com-vishalorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d> (Accessed 4 January 2021).
12. Medium (2018), "How Ai Will Help You Avoid Incompetence", available at: <https://medium.com/causys/how-ai-will-help-you-avoid-incompetence-7d8b8e212014> (Accessed 4 January 2021).

Стаття надійшла до редакції 19.01.2021 р.