ВІСНИК ПРИАЗОВСЬКОГО ДЕРЖАВНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ
2025р. Серія: Технічні науки Вип. 52
p-ISSN: 2225-6733; e-ISSN: 2519-271X

УДК 004.85:331.108 DOI: 10.31498/2225-6733.52.2025.350986

МЕТОДИ МАШИННОГО НАВЧАННЯ ЯК ІНСТРУМЕНТ АНАЛІЗУ ТЕМПІВ КАР'ЄРНОГО РОЗВИТКУ СПІВРОБІТНИКІВ ІТ-КОМПАНІЇ

Сиротенко О.Г. *магістр, Харківський національний університет радіоелектроніки, м. Харків, ORCID: https://orcid.org/0009-0009-9614-9056, e-mail: olena.syrotenko@nure.ua;*

Петрова Р.В. *канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, м. Харків, ORCID: https://orcid.org/0000-0001-5886-8943, e-mail: roksana.petrova@nure.ua;*

Морозова А.І. *канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, м. Харків, ORCID: https://orcid.org/0000-0002-7082-4115, e-mail: anna.morozova@nure.ua*

У статті детально досліджується використання методів машинного навчання в задачах управління людськими ресурсами в ІТ-компаніях. В умовах високої конкуренції за кваліфіковані кадри, розуміння та прогнозування кар'єрного розвитку співробітників стає ключовим фактором для їх утримання та мотивації. Використання інструментів машинного навчання дозволяє виявити приховані закономірності у великих обсягах HR-даних і сформувати об'єктивну та ефективну картину розвитку персоналу. Метою роботи є аналіз можливостей використання методів машинного навчання для визначення динаміки кар'єрного розвитку співробітників ІТ-компанії. Завданнями роботи є формування практичного алгоритму виявлення груп фахівців за ефективністю їхнього розвитку та оцінка чинників, що впливають на темпи кар'єрного зростання. У процесі дослідження використано комплекс методів інтелектуального аналізу даних та машинного навчання, що включають методи кластеризації (метод K-Means, метод головних компонент), методи оцінки якості кластеризації (силуетний коефіцієнт), методи регресійного аналізу (лінійна регресія, B-сплайн регресія) та методи оцінки якості регресійного аналізу (метод кореня середньоквадратичної помилки, коефіцієнт детермінації). У роботі виконано попередню обробку експериментальних даних, сформульовано ознаки ефективності зростання та застосовано методи кластеризації для виявлення схожих груп співробітників за сформульованими ознаками. Також досліджено взаємозв'язок між чинником «кількість проектів» та темпами кар'єрного розвитку із подальшою оцінкою його впливу з використанням методів регресійного аналізу. У результаті дослідження виокремлено три основні групи співробітників: з високим темпом зростання, без зростання та з нестабільним зростанням. Виявлено статистично значущий вплив кількості проектів на темпи кар'єрного розвитку та нелінійну залежність між цими ознаками. Практична значущість роботи полягає в тому, що розроблені моделі та підходи можуть бути інтегровані в HR-системи ІТ-компаній для об'єктивної оцінки персоналу, формування кадрового резерву та розробки персоналізованих планів розвитку.

Ключові слова: *ІТ-компанія, інтелектуальний аналіз даних, інформаційні технології, кар'єрний розвиток, кластеризація, машинне навчання, оптимізація, регресійний аналіз.*

Постановка проблеми

Інформаційні технології є однією з найдинамічніших галузей сучасної економіки, що характеризується високими темпами розвитку та постійними технологічними змінами.

ІТ-компанії функціонують у висококонкурентному середовищі, де ключовим ресурсом є кваліфіковані фахівці. На відміну від традиційних галузей, де основу створюють матеріальні активи та технологічне обладнання, в ІТ-сфері інтелектуальний потенціал співробітників безпосередньо впливає на якість продукту, швидкість інновацій та здатність адаптуватися до ринкових змін.

Високий рівень мобільності ІТ-спеціалістів перетворює утримання співробітників на стратегічний пріоритет. Втрата ключового співробітника може призвести не лише до прямих фінансових витрат на заміщення позиції, але й до втрати критичних знань та порушення проектних термінів.

Традиційні підходи до управління персоналом, розроблені для стабільних галузей з передбачуваними кар'єрними шляхами, виявляються неефективними [1] в динамічному ІТ-середовищі, де вимоги до компетенцій співробітника змінюються відповідно до темпів розвитку інформаційних технологій.

ІТ-компанії накопичують значну кількість історичних даних про кар'єру співробітників, включаючи їхній досвід, технічні навички, кількість проектів, в яких брав участь співробітник, його роль у проекті, тривалість роботи на проекті, тривалість перебування на посаді, рівень заробітної плати, географічне розташування, рівень освіти, сертифікати, оцінка ефективності з боку команди та менеджерів тощо.

Це створює передумови для застосування інтелектуальних методів аналізу даних та машинного навчання, які здатні виявляти приховані закономірності та тренди, зокрема для прогнозування персонального кар'єрного розвитку співробітника та визначення загальних чинників впливу на темпи кар'єрного розвитку

Copyright © 2025, Authors. This is an open access article under the Creative Commons CC BY license

p-ISSN: 2225-6733; e-ISSN: 2519-271X

співробітників, покриваючи економічні та стратегічні цілі ІТ-компанії.

Економічні переваги інтелектуального аналізу кар'єрної динаміки співробітників для ІТ-компанії включають:

– зниження витрат на підбір співробітників – ефективне планування кар'єрного розвитку існуючих співробітників зменшує витрати на залучення нового ІТ-спеціаліста;

– зменшення плинності кадрів – наявність персоналізованих планів розвитку позитивно впливає на утримання співробітників в ІТ-компаніях, що впливає на зниження витрат, пов'язаних з плинністю персоналу;

– підвищення продуктивності – співробітники із чіткими перспективами кар'єрного зростання демонструють вищі показники продуктивності та залученості в роботу [2], що впливає на збільшення кількості клієнтів ІТ-компанії.

Стратегічні переваги інтелектуального аналізу кар'єрної динаміки співробітників для ІТ-компанії включають:

– формування кадрового резерву – інтелектуальний аналіз дозволяє виявляти співробітників з високим потенціалом до ключових позицій та заздалегідь готувати їх до них, забезпечуючи безперервність управління;

– підвищення конкурентоспроможності – ІТ-компанії, що ефективно розвивають свій персонал, отримують конкурентні переваги за рахунок вищої кваліфікації команди та здатності швидше впроваджувати інновації.

Актуальність проблеми полягає в необхідності переходу від суб'єктивних оцінок до об'єктивного, керованого даними підходу в управлінні кар'єрою. Використання машинного навчання дозволяє не лише описати поточну динаміку, але й сформувати підґрунтя для побудови ефективних систем мотивації, розвитку та утримання фахівців.

Аналіз останніх досліджень та публікацій

У дослідженнях HR-асоціації SHRM (Society for Human Resource Management) підкреслюється, що саме можливості кар'єрного розвитку та професійного навчання є ключовим фактором утримання фахівців у технологічному секторі, часто випереджаючи рівень фінансової компенсації [3].

У роботі [4] автори доводять, що неефективне управління кар'єрою та відсутність прозорих, персоналізованих шляхів розвитку є прямим каталізатором високої плинності кадрів. Це, в свою чергу, завдає компаніям значних фінансових збитків, пов'язаних із пошуком, наймом та адаптацією нових спеціалістів.

Ці виклики стимулювали розвиток нового напрямку «People Analytics» [5], що ґрунтується на зборі та інтелектуальному аналізі великих обсягів HR-даних для прийняття обґрунтованих управлінських рішень.

Значна частина сучасних наукових праць [6-8] у цій сфері зосереджена на вирішенні задачі прогнозування підвищення (promotion prediction). У цих роботах автори розглядають підвищення як бінарну класифікаційну подію.

У статті [6] автори Ch.Dayakar та P.Pavani пропонують методологію прогнозування, що базується на таких алгоритмах, як Support Vector Machine, Artificial Neural Network та Random Forest. Їхня модель прогнозує, чи буде співробітник підвищений, на основі таких ознак, як вислуга років, рівень продуктивності та бали за пройдене навчання. Ключовим внеском їхньої роботи є вирішення проблеми незбалансованих наборів даних за допомогою техніки ROS (Random Over-Sampling), що дозволило моделі Random Forest досягти 98% точності.

У дослідженні [7] автори порівнюють ефективність Logistic Regression, Random Forest, Gradient Boosting та Decision Tree. Вони використовують набір даних, що включає вік, кількість тренінгів та стаж роботи. У їхньому випадку найвищу точність у 86% показала логістична регресія.

Автори в роботі [8] розширюють цей список, додаючи до порівняння AdaBoost та XGBoost, і в їхньому експерименті найкращі результати демонструє модифікований підхід AdaBoost.

Група авторів [9] пропонує більш комплексний підхід, що виходить за межі простого прогнозування, – рекомендаційні системи для кар'єрних шляхів. Вони критикують застарілу модель «кар'єрної драбини» та описують архітектуру системи «CareerRec», яка використовує ML (зокрема, XGBoost) не для прогнозування, а для рекомендації персоналізованих, нелінійних кар'єрних шляхів на основі профілю навичок фахівця.

У статті [10] автори пропонують метод кластеризації послідовностей (sequence clustering) на основі вбудовування (embedding) для аналізу профілів ІТ-фахівців. Їхній підхід дозволяє виявити неявні зв'язки між посадами та сформувати типові кар'єрні шляхи для кожної групи.

Огляд літератури показує, що більшість праць фокусується на прогнозуванні факту підвищення аналізу структури кар'єрних шляхів, натомість, недостатньо дослідженими залишаються питання аналізу темпів кар'єрного розвитку як динамічної величини, а також класифікація співробітників за шаблонами їхнього зростання.

Мета статті

Метою роботи є формулювання варіантів практичного використання методів машинного навчання для аналізу темпів кар'єрного розвитку співробітників ІТ-компанії, включаючи задачі ідентифікації груп співробітників за ефективністю їхнього кар'єрного розвитку та аналізу ступеня кореляції і впливу професійних

p-ISSN: 2225-6733; e-ISSN: 2519-271X

ознак співробітника на ефективність його кар'єрного розвитку.

Матеріали та методи

Об'єктом дослідження є процес кар'єрного розвитку співробітників в IT-компаніях.

Предметом дослідження є методи машинного навчання аналізу динаміки кар'єрного розвитку співробітників IT-компанії.

Для валідації запропонованих в роботі підходів використання методів машинного навчання для аналізу темпів кар'єрного розвитку використано експериментальний набір даних, що відображає типові, анонімізовані HR-записи IT-компанії за період з січня 2023 по лютий 2025 року.

Набір даних складається з двох основних джерел:

– щомісячні звіти про співробітників. Представлені у вигляду набору файлів відповідно до кожного місяця. Кожен запис у цих файлах містить ключові атрибути, що описують стан співробітника на певний місяць, зокрема: ідентифікатор співробітника, команда, посадовий рівень та дата найму;

– дані про проектну діяльність. Цей набір містить деталі про участь співробітників у проектах, включаючи ідентифікатор співробітника, ідентифікатор проекту, дату початку роботи на проекті, дату завершення роботи на проекті та відсоток завантаження.

Методи дослідження включають методи кластеризації (метод K-Means, метод головних компонент), методи оцінки якості кластеризації (силуетний коефіцієнт), методи регресійного аналізу (лінійна регресія, B-сплайн регресія) та методи оцінки якості регресійного аналізу (метод кореня середньоквадратичної помилки, коефіцієнт детермінації).

Метод кластеризації K-Means полягає в розділенні n спостережень на k кластерів, так щоб кожне спостереження належало до кластера з найближчим до нього середнім значенням [11]. Кластеризація є процесом ітеративним і включає визначення центроїдів та оцінку відстані між кожним спостереженням та присвоєним йому центроїдом.

Центроїд представляє середнє координат всіх точок кластера за кожною ознакою відповідно до формули (1).

$$m_i = \frac{1}{|S_i|}\sum_{x \epsilon S_i} x, \quad (1)$$

де S_i – множина всіх точок, що належать i-ому кластеру;

$|S_i|$ – кількість точок в i-ому кластері;

x – окрема точка, що належить i-ому кластеру.

Відстань між кожним спостереженням та центроїдом визначається як квадратична Евклідова відстань, відповідно до якої спостереження присвоюється цетроїду з найменшою квадратичною Евклідовою відстанню, як наведено в формулі (2).

$$S_i^{(t)} = \{x_p : \left|\left|x_p - m_i^{(t)}\right|\right|^2 \leq \left|\left|x_p - m_j^{(t)}\right|\right|^2 \ \forall j, 1 \leq j \leq k\}, \quad (2)$$

де $S_i^{(t)}$ – i-й кластер на ітерації t;

x_p – p-та точка даних у наборі даних;

$m_i^{(t)}$ – центроїд i-го кластера на ітерації t;

$m_j^{(t)}$ – центроїд j-го кластера на ітерації t;

k – загальна кількість кластерів.

Для візуалізації результатів кластерного аналізу в двовимірному просторі використано метод головних компонент (PCA), який призначений для зменшення розмірності даних з мінімальними втратами. Метод здійснює лінійне перетворення вихідних змінних у нові, незалежні одна від одної головні компоненти, які впорядковуються за спаданням дисперсії. Перша головна компонента відображає найбільше розсіювання у вибірці, остання – найменше розсіювання. Формула (3) відображає процес проектування даних на головні компоненти.

$$Z = X \cdot W, \quad (3)$$

де X – центрована матриця вихідних даних розміром $n \times p$, де n – кількість спостережень, p – кількість ознак;

W – матриця власних векторів розміром $p \times k$, де k – кількість головних компонент.

Для визначення залежностей між ознаками використано методи лінійної регресії та сплайн-регресії на основі B-сплайну.

Лінійна регресія моделює лінійну залежність між скалярною змінною y та векторною змінною X. Модель лінійної регресії наведено у формулі (4) [12].

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon, \quad (4)$$

де y – залежна пояснювальна змінна;

β_0 – вільний член;

$(x_1,\ldots,x_n)$ – незалежні пояснювальні змінні;

β_n – коефіцієнт моделі;

ε – випадкова похибка, розподіл якої в загальному випадку залежить від незалежних змінних, але математичне сподівання якої дорівнює нулю;

n – кількість ознак (незалежних змінних).

Коефіцієнти підбираються так, щоб мінімізувати суму квадратів відхилень між фактичними та передбаченими значеннями, реалізуючи метод найменших квадратів. Лінійну регресію обрано як перший крок для виявлення глобальних тенденцій перед застосуванням більш гнучкої сплайн-моделі.

Сплайн регресія дозволяє моделювати нелінійні залежності між змінними, розбиваючи простір ознак на окремі ділянки та апроксимуючи їх різними функціями. У цьому підході простір незалежної змінної розбивається на окремі інтервали за допомогою вузлів, а на кожному інтервалі функція апроксимується

p-ISSN: 2225-6733; e-ISSN: 2519-271X

поліномом заданого ступеня [13]. Модель B-сплайн регресії наведено у формулі (5).

$$y_i = \beta_1 B_1(x_1) + \cdots + \beta_k B_k(x_n) + \varepsilon_i, \quad (5)$$

де $B_k(x)$ – базисна функція B-сплайну;

k – кількість базисних функцій.

На відміну від лінійної регресії, де вся модель є прямою або гіперплощиною, сплайнова модель дозволяє згинатися в різних точках, краще підлаштовуючись до трендів у даних та забезпечуючи неперервність похідних у вузлах, що дозволяє відображати поведінку залежності в різних діапазонах без втрати гладкості.

Виклад основного матеріалу

На першому етапі дослідження виконано попередню обробку експериментальних даних. Вихідні матеріали складалися з двох анонімізованих наборів: щомісячних звітів про стан співробітників та окремого файлу з даними про їхню проектну діяльність . Після завантаження, консолідації щомісячних файлів в єдину структуру та видалення дублікатів проведено видалення повних дублікатів.

Перед застосуванням методів машинного навчання важливо виконати аудит повноти даних про співробітників, оскільки пропущені значення можуть призвести до викривлення результатів або неможливості застосування певних методів [14].

При виявленні пропущених значень критично важливо оцінити їхній відсоток та загальний вплив на вибірку. Поширеною помилкою є повне видалення рядків з пропусками (listwise deletion), оскільки це може суттєво зменшити обсяг вибірки та, як наслідок, статистичну потужність подальшого аналізу. Тому, при некритичному відсотку пропусків, перевага надається методам заповнення (imputation).

У даному дослідженні, слідуючи цьому підходу, аналіз виявив некритичні пропуски у полі «команда», що становить 10,43%. Щоб зберегти обсяг вибірки, застосовано заповнення маркерами «No Team». В результаті отримано чистий, повний набір даних , валідований на відсутність пропусків, що є необхідною умовою для коректного застосування методів машинного навчання на наступних етапах.

Перед початком аналізу динаміки кар'єрного розвитку додатково введено ієрархію посад, де кожному кар'єрному рівню присвоєно числовий ранг, як наведено в табл. 1. Використання такого словника забезпечить коректне сортування даних при подальшому аналізі та візуалізації, оскільки стандартне алфавітне сортування не відображає реальної ієрархії посад в організації.

Для виявлення груп фахівців за ефективністю їхнього розвитку застосовано метод кластеризації k-середніх.

Таблиця 1

Словник ієрархії посад в IT-компанії

Посада	Рівень
Trainee	1
Junior	2
Middle	3
Senior	4
Lead	5
Expert	6
Associate Director	7
Director	8

В якості вектора ознак ефективності кар'єрного зростання співробітника V_i пропонується використати вектор (6) наступних ознак: індекс зрілості, кількість позицій у кар'єрній траєкторії співробітника в компанії, індекс стабільності зростання та факт підвищення.

$$V_i = (MaturityIndex_i, Levels_i, Stability_i, Grew_i). \quad (6)$$

Завданням методу k-середніх є розділення множини всіх векторів $V = \{V_1, \ldots, V_n\}$ на k кластерів, мінімізуючи суму квадратів відстаней від кожного вектора V_i до центроїда його кластера.

Пропонується визначати індекс зрілості відповідно до (7), що дозволить інтегрувати поточний рівень позиції та загальний стаж роботи через зважену комбінацію нормалізованих значень, дозволяючи об'єктивно порівняти співробітників.

$$MaturityIndex = 10 \cdot (\alpha \cdot L + (1 - \alpha) \cdot E), \quad (7)$$

де L – ієрархічне значення поточної позиції співробітника;

E – стаж роботи співробітника в днях;

α – ваговий коефіцієнт важливості рівня посади.

Параметри ієрархічного значення поточної позиції співробітника та стажу роботи мають бути попередньо нормалізованими до діапазону від 0 до 1, з метою забезпечення рівнозначного внеску обох компонентів в індекс зрілості.

Для вагового коефіцієнту α пропонується обрати значення відповідно до суб'єктивної оцінки важливості рівня посади над стажем роботи. В даному експерименті встановлено значення 0,2, що означає, що поточний рівень має вагу 20% в індексі зрілості, тоді як стаж роботи 80%.

Фінальний розрахунок індексу зрілості включає множення зваженої суми нормалізованих значень на коефіцієнт 10 для отримання шкали від 0 до 10.

Кількість позицій у кар'єрній траєкторії співробітника в компанії відповідає розміру множини усіх рівнів посад, які співробітник займав (8):

$$Levels = |\{pos_i \mid (pos_i, _) \in P\}|, \quad (8)$$

p-ISSN: 2225-6733; e-ISSN: 2519-271X

де P – кар'єрна траєкторія співробітника, представлена як кортеж всіх його послідовних етапів роботи, $P = [step_1, step_2, \ldots, step_n]$, $step_i = (pos_i, days_i)$;

pos_i – рівень посади.

Індекс стабільності зростання дозволяє оцінити ефективність кар'єрних переходів: високе значення вказує на послідовне просування вгору по ієрархії, тоді як низьке значення свідчить про повернення на попередні рівні. Пропонується визначати індекс стабільності зростання як відношення кількості унікальних рівнів до загальної кількості переходів (9):

$$Stability = \frac{Levels}{Transitions}, \quad (9)$$

де *Levels* – кількість позицій у кар'єрній траєкторії співробітнкиа;

Transitions – загальна кількість етапів у кар'єрній траєкторії співробітника, $Transitions = |P|$.

Факт підвищення визначено у вигляді бінарної змінної, яка приймає значення True, якщо співробітник має хоча б 1 підвищення, та False, якщо співробітник залишається на одній позиції (10):

$$Grew = \begin{cases} 1, & Levels > 1 \\ 0, & Levels \leq 1 \end{cases} \quad (10)$$

Кількість кластерів k пропонується визначати експериментально відповідно до оцінки якості кластеризації з використанням силуетного коефіцієнту (11):

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),\ b(i)\}}, \quad (11)$$

де *a(i)* – середня відстань від *i*-го об'єкта до всіх інших об'єктів у тому ж кластері;

b(i) – найменша середня відстань від *i*-го об'єкта до всіх об'єктів найближчого кластера.

Силуетний коефіцієнт дозволяє кількісно визначити, наскільки добре кожен об'єкт належить до свого кластеру порівняно з іншими. У даному дослідженні оптимальним значенням обрано $k = 3$, при якому силуетний коефіцієнт становить найбільше значення 0,7, що вказує на якість кластеризації вище середньої.

В результаті кластеризації визначено 3 кластери із середніми значеннями ознак ефективності кар'єрного зростання кожного кластеру, як наведено в табл. 2.

Таблиця 2

Кластери співробітників за ефективністю кар'єрного розвитку

Номер кластеру	Індекс зрілості (*MaturityIndex*)	Кількість унікальних позицій (*Levels*)	Індекс стабільності кар'єрного зростання (*Stability*)	Факт підвищення (*Grew*)	Категорія кластеру
1	3,57	2,28	1	1	Високоефективні
2	1,59	1	1	0	Потенційно застійні
3	4,88	2,5	0,75	1	Нестабільні

На основі отриманих результатів можна класифікувати кожен кластер наступним чином:

– кластер 1 містить дані про співробітників з високими показниками зростання, стабільності та середнім індексом зрілості, що свідчить про ефективне зростання співробітників в цьому кластері;

– кластер 2 містить дані про співробітників, які не мали кар'єрного зростання, що свідчить про відсутність кар'єрного розвитку співробітників в цьому кластері;

– кластер 3 включає співробітників із нестабільною динамікою зростання.

Візуалізація кластерів за допомогою PCA (рис. 1) чітко демонструє розділення між групами високоефективних співробітників («High Growth») та малоефективних («No Growth») співробітників. Натомість кластер нестабільних співробітників («Unstable Growth») представлений лише кількома спостереженнями з великою відстанню від основної маси точок, що визначає аномальний характер даних в цьому кластері, які сильно відрізняються від більшості.

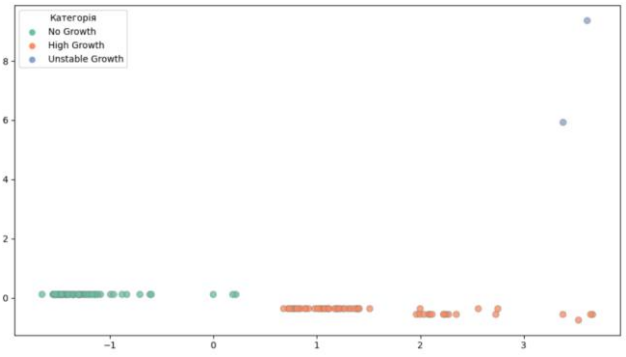

Рис. 1 – Візуалізація кластеризації співробітників за ознаками ефективності кар'єрного зростання

Запропонована кластеризація дозволяє IT-компанії застосовувати персоналізовані стратегії управління талантами. Замість того, щоб аналізувати лише факт підвищення, цей метод дає якісну оцінку шаблону розвитку. Це дозволяє ідентифікувати не лише очевидних лідерів, але й приховані групи ризику, які потребують проактивного втручання.

Наприклад, фахівці з групи високоефективних співробітників можуть бути пріоритезовані для

p-ISSN: 2225-6733; e-ISSN: 2519-271X

включення до кадрового резерву. Окрім того, аналіз їхніх успішних траєкторій дає змогу виявити найефективніші кар'єрні шляхи та використовувати їх як основу для корпоративних програм менторства.

На противагу, кластер малоефективних співробітників є ключовим індикатором ризику стагнації та плинності кадрів. Замість того, щоб реагувати на звільнення постфактум, компанія може аналізувати причини відсутності зростання в цій групі та впроваджувати цільові програми для їх утримання.

Особливу аналітичну цінність становить кластер нестабільних працівників. Хоча ці співробітники можуть мати високий індекс зрілості, низький показник стабільності вказує на неефективні «горизонтальні» або «зворотні» кар'єрні переміщення. Для HR-менеджерів це є прямим сигналом для аналізу причин такої нестабільності – проблема в плануванні проектів, відсутність чіткого кар'єрного бачення тощо.

Для виявлення та оцінки чинників, що впливають на темпи кар'єрного зростання пропонується використати такі методи регресійного аналізу, як лінійна регресія та B-сплайн регресія.

Для оцінки якості регресійної моделі обрано метод кореня середньоквадратичної помилки (RMSE) коефіцієнт детермінації (R^2) [15].

Метод кореня середньоквадратичної помилки є квадратним коренем із середнього значення квадратів різниць між прогнозованими та спостережуваними результатами, як наведено в формулі (12):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad (12)$$

де n – загальна кількість спостережень;

y_i – фактичне значення залежної змінної для i-го спостереження;

$\hat{y}_i$ – передбачене значення залежної змінної для i-го спостереження.

Чим менше отримана метрика, тим краща здатність моделі до точного прогнозування. І навпаки, більший результат означає більшу розбіжність між прогнозованими та фактичними результатами. На відміну від методу середньої абсолютної похибки, метод кореня середньоквадратичної помилки більш суворо карає великі помилки за рахунок піднесення до квадрату.

Коефіцієнт детермінації (R^2) оцінює, яку частку дисперсії цільової змінної пояснює модель та обчислюється відповідно до формули (13):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}, \quad (13)$$

де $\bar{y}_i$ – середнє фактичне значення залежної змінної.

Коефіцієнт детермінації показує, наскільки краще модель представляє дані в порівнянні з середнім значенням: значення 1 означає ідеальне передбачення, 0 вказує, що модель не краща за просте середнє значення, а від'ємне значення свідчить про те, що модель є гіршою за середнє значення.

Використання лінійної регресії дозволяє виявити глобальні тенденції між обраною ознакою й цільовою та прийняти рішення про доцільність застосування більш гнучкої сплайн-моделі.

Цільовою ознакою є темпи кар'єрного розвитку співробітника, яка описується формулою (14):

$$Growth\ Rate = \frac{L_{end} - L_{start}}{D/365}, \quad (14)$$

де L_{start}, L_{end} – ієрархічні рівні початкової та поточної позицій співробітника;

$D/365$ – загальна кількість років на посадах.

В якості досліджуваної ознаки в даній роботі розглядається ознака «кількість проектів». Таким чином, ми формуємо гіпотези:

– H_0: кількість проектів не має статистично значущого впливу на темпи кар'єрного зростання;

– H_1: кількість проектів має статистично значущий вплив на темпи кар'єрного зростання.

Побудовану лінійну регресійну модель впливу кількості проектів на темпи кар'єрного розвитку наведено на рис. 2.

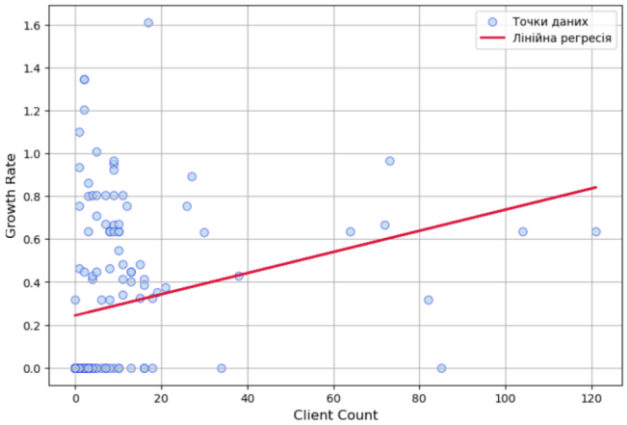

Рис. 2 – Лінійна регресійна модель впливу кількості проектів на темпи кар'єрного розвитку

Для лінійної регресійної моделі корінь середньоквадратичної помилки становить 0,13; коефіцієнт детермінації дорівнює 0,6.

Статистичний тест підтвердив, що кількість клієнтів має значущий вплив (p = 0,003 < 0,005), проте лінійна регресійна модель пояснює близько 6% дисперсії в темпі зростання, що є слабким результатом для практичної інтерпретації.

Таким чином, лінійна модель не є ефективною моделлю для точного прогнозування темпу зростання на основі кількості клієнтів.

У випадку неефективності використання лінійних методів для опису залежностей між ознаками пропонується використати регресійну модель на основі B-

p-ISSN: 2225-6733; e-ISSN: 2519-271X

сплайнів. Кількість сплайнів підбирається експериментально, орієнтуючи на результати методів оцінки якості регресії.

В даному дослідженні використано регресійну модель на основі 4 B-сплайнів (рис. 3).

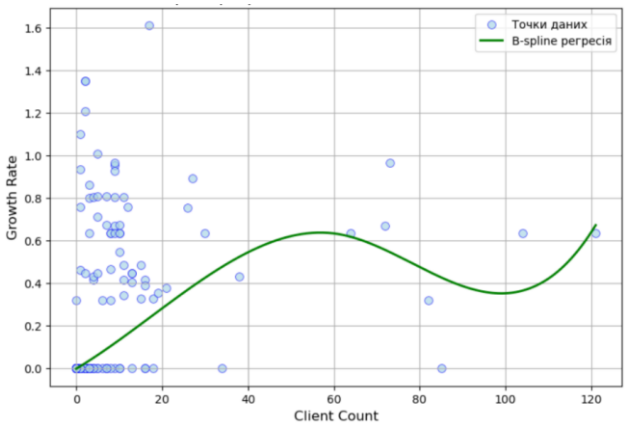

Рис. 3 – Сплайн регресійна модель впливу кількості проектів на темпи кар'єрного розвитку

Корінь середньоквадратичної помилки сплайн регресійної моделі становить 0,12, а коефіцієнт детермінації дорівнює 0,19.

Можемо визначити, що побудована сплайн регресійна модель є кращою за лінійну, оскільки має менше помилок та вищий коефіцієнт детермінації, проте модель пояснює близько 19% дисперсії, що все ще є невисоким результатом для точного прогнозування, а отже зв'язок між обраними ознаками слабкий.

Незважаючи на це, модель дозволяє визначити, що до 40 клієнтів темпи зростання співробітників були стабільними, а після 40 почали коливатися, а отже занадто високе навантаження клієнтами не гарантує швидкого кар'єрного зростання співробітника.

Застосування запропонованого комплексу регресійних методів (лінійної та B-сплайн регресії) може бути інтегроване у системи «People Analytics» як ефективний інструмент для діагностики та оцінки чинників, що впливають на темпи кар'єрного зростання. Цей підхід дозволяє вийти за межі простих кореляцій та отримати глибоке розуміння характеру взаємозв'язків між ознаками, що дозволить оптимізувати управлінські рішення та уникнути умов, що призводять до професійної стагнації чи вигорання співробітників.

Висновки

У ході дослідження сформульовано та апробовано комплекс практичних методів машинного навчання для аналізу темпів кар'єрного розвитку співробітників IT-компанії.

На відміну від більшості існуючих підходів, що фокусуються на бінарному прогнозуванні факту підвищення, дана робота вирішує два ключових аналітичних завдання:

– сформовано практичний алгоритм виявлення груп фахівців за ефективністю їхнього розвитку. Запропоновано набір ознак, який включає факт підвищення, індекс зрілості, кількість позицій та індекс стабільності кар'єрного зростання. Застосування методу кластеризації K-Means на основі цього вектора дозволило успішно виокремити три якісно різні групи співробітників: «Високоефективні», «Потенційно застійні» та «Нестабільні»;

– сформовано практичній підхід оцінки чинників, що впливають на темпи кар'єрного зростання. Для цього введено динамічну метрику «Growth Rate», що кількісно вимірює швидкість просування кар'єрними сходами. Двохетапний регресійний аналіз на прикладі ознаки «кількість проектів» дозволив дослідити її нелінійний вплив на темпи кар'єрного розвитку та визначити, що занадто високе навантаження клієнтами не гарантує швидкого кар'єрного зростання.

Запропоновані методи є готовими інструментами для інтеграції в системи «People Analytics». Вони дозволять HR-департаментам перейти від суб'єктивних оцінок до об'єктивної діагностики персоналу, ідентифікувати групи ризику стагнації та оптимізувати управлінські рішення.

Перелік використаних джерел

[1] Nastase C., Adomnitei A., Apetri A. Strategic Human Resource Management in the Digital Era: Technology, Transformation, and Sustainable Advantage. *Merits*. 2025. Vol. 5, no. 4. Article 23. DOI: https://doi.org/10.3390/merits5040023.

[2] The Effect of Training and Career Development on Company Performance: A Systematic Literature Review / N. T. Fabian et al. *International Journal of Humanities, Law, and Politics*. 2024. Vol. 2, no. 1. Pp. 19-25. DOI: https://doi.org/10.46336/ijhlp.v2i1.57.

[3] 2025 Employee Benefits Survey. URL: https://www.shrm.org/content/dam/en/shrm/topics-tools/research/employee-benefits/2025_annual_benefits_survey_executive_summary.pdf (дата звернення: 10.10.2025).

[4] Ullah S., Azeem S. Effect of Employee Retention in the IT Industry. *Global Management Sciences Review*. 2024. Vol. IX, no. II. Pp. 83-89. DOI: https://doi.org/10.31703/gmsr.2024(ix-ii).08.

[5] Han Wuen C. The Application of People Analytics to Enhance Decision-Making in Human Resource Management. *Ambidextrous Journal of Innovation Efficiency and Technology in Organization*. 2025. Vol. 2, no. 02. Pp. 88-96. DOI: https://doi.org/10.61536/ambidextrous.v2i02.179.

[6] Dayakar Ch., Pavani P. Machine Learning Based Employee Promotion Predition. *Journal of Engineering Sciences*. 2023. Vol 14, no. 08. Pp. 417-423.

[7] Ansari N., Vora N. Employee Promotion Evaluation and Prediction using Machine Learning. *Journal of*

Information Technology and Digital World. 2024. Vol. 6, no. 4. Pp. 317-332. DOI: https://doi.org/10.36548/jitdw.2024.4.001.

[8] Employee Promotion Prediction Using Improved AdaBoost Machine Learning Approach / M. A. Jafor et al. *AIUB Journal of Science and Engineering (AJSE)*. 2023. Vol. 22, no. 3. Pp. 258-266. DOI: https://doi.org/10.53799/ajse.v22i3.781.

[9] A Machine Learning Approach to Career Path Choice for Information Technology Graduates / H. Al-Dossari et al. *Engineering, Technology & Applied Science Research*. 2020. Vol. 10, no. 6. Pp. 6589-6596. DOI: https://doi.org/10.48084/etasr.3821.

[10] Zhong H., Liu C., Wu C. Uncovering IT Career Path Patterns with Job Embedding-based Sequence Clustering. *ACM Transactions on Management Information Systems*. 2025. Vol. 16, iss. 2. Pp. 1-32. DOI: https://doi.org/10.1145/3712705.

[11] Hollander M., Wolfe D., Chicken E. Nonparametric Statistical Methods. John Wiley & Sons, Inc., 2014. 848 p. DOI: https://doi.org/10.1002/9781119196037.

[12] Montgomery D. C., Peck E. A., Vining G. G. Introduction to Linear Regression Analysis. 6th ed. John Wiley & Sons, Inc., 2021. 704 p.

[13] Bowman A., Azzalini A. Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations. Oxford University Press, 1997. 204 p. DOI: https://doi.org/10.1093/oso/9780198523963.001.0001.

[14] Han J., Kamber M., Pei J. Data Mining Concepts & Techniques. 3rd ed. Morgan Kaufmann Publishers, 2011. 744 p.

[15] Harrel F. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer, 2001. 600 p. DOI: https://doi.org/10.1007/978-1-4757-3462-1.

MACHINE LEARNING METHODS AS A TOOL FOR ANALYSING THE CAREER DEVELOPMENT PACE OF IT COMPANY EMPLOYEES

Syrotenko O.G. *Master's degree candidate, Kharkiv National University of Radio Electronics, Kharkiv, ORCID: https://orcid.org/0009-0009-9614-9056, e-mail: olena.syrotenko@nure.ua;*

Petrova R.V. *PhD (Engineering), associate professor, Kharkiv National University of Radio Electronics, Kharkiv, ORCID: https://orcid.org/0000-0001-5886-8943, e-mail: roksana.petrova@nure.ua;*

Morozova A.I. *PhD (Engineering), associate professor, Kharkiv National University of Radio Electronics, Kharkiv, ORCID: https://orcid.org/0000-0002-7082-4115, e-mail: anna.morozova@nure.ua*

The article examines in detail the use of machine learning methods in human resource management tasks in IT companies. In a highly competitive environment for qualified employees, understanding and predicting employee career development becomes a key factor in retaining and motivating them. The use of machine learning tools allows us to identify hidden patterns in large volumes of HR data and form an objective and effective picture of staff development. The purpose of this work is to analyse the possibilities of using machine learning methods to determine the dynamics of career development of IT company employees. The objectives of this work are to develop a practical algorithm for identifying groups of specialists based on the effectiveness of their development and to assess the factors that influence the pace of career growth. The research uses a set of methods of data mining and machine learning, including clustering methods (K-Means method, principal component method), methods for assessing the quality of clustering (silhouette coefficient), methods of regression analysis (linear regression, B-spline regression), and methods for assessing the quality of regression analysis (root mean square error method, coefficient of determination). The paper performs preliminary processing of experimental data, formulates indicators of growth efficiency, and applies clustering methods to identify similar groups of employees based on the formulated indicators. The relationship between the factor «number of projects» and career development rates was also studied, with a subsequent assessment of its impact using regression analysis methods. As a result of the study, three main groups of employees were identified: those with high growth rates, those without growth, and those with unstable growth. A statistically significant impact of the number of projects on the pace of career development and a nonlinear relationship between these characteristics were identified. The practical significance of the work lies in the fact that the developed models and approaches can be integrated into the HR systems of IT companies for objective personnel assessment, the formation of a personnel reserve, and the development of personalized development plans.

Keywords: *IT company, data mining, information technology, career development, clustering, machine learning, optimization, regression analysis.*

References

[1] C. Nastase, A. Adomnitei, and A. Apetri, "Strategic Human Resource Management in the Digital Era: Technology, Transformation, and Sustainable Advantage," *Merits*, vol. 5, no. 4, p. 23, 2025. **doi: 10.3390/merits5040023**.

[2] N. T. Fabian et al., "The Effect of Training and Career Development on Company Performance: A Systematic Literature Review," *International Journal of Humanities, Law, and Politics*, vol. 2, no. 1, pp. 19-25, 2024. **doi: 10.46336/ijhlp.v2i1.57**.

[3] 2025 Employee Benefits Survey. [Online]. Available: https://www.shrm.org/content/dam/en/shrm/topics-tools/research/employee-benefits/2025_annual_benefits_survey_executive_summary.pdf. Accessed on: October 10, 2025.

[4] S. Ullah, and S. Azeem, "Effect of Employee Retention in the IT Industry," *Global Management Sciences Review*, vol. IX, no. II, pp. 83-89, 2024. **doi: 10.31703/gmsr.2024(ix-ii).08**.

[5] C. Han Wuen, "The Application of People Analytics to Enhance Decision-Making in Human Resource Management," *Ambidextrous Journal of Innovation Efficiency and Technology in Organization*, vol. 2, no. 02, pp. 88-96, 2025. **doi: 10.61536/ambidextrous.v2i02.179**.

[6] Ch. Dayakar, and P. Pavani, "Machine Learning Based Employee Promotion Predition," *Journal of Engineering Sciences*, vol. 14, no. 08, pp. 417-423, 2023.

[7] N. Ansari, and N. Vora, "Employee Promotion Evaluation and Prediction using Machine Learning," *Journal of Information Technology and Digital World*, vol. 6, no. 4, pp. 317-332, 2024. **doi: 10.36548/jitdw.2024.4.001**.

[8] M.A. Jafor, M.A.H. Wadud, K. Nur, and M.M. Rahman, "Employee Promotion Prediction Using Improved AdaBoost Machine Learning Approach," *AIUB Journal of Science and Engineering (AJSE)*, vol. 22, no. 3, pp. 258-266, 2023. **doi: 10.53799/ajse.v22i3.781**.

[9] H. Al-Dossari, F.A. Nughaymish, Z. Al-Qahtani, M. Alkahlifah, and A. Alqahtani, "A Machine Learning Approach to Career Path Choice for Information Technology Graduates," *Engineering, Technology & Applied Science Research*, vol. 10, no. 6, pp. 6589-6596, 2020. doi: 10.48084/etasr.3821.

[10] H. Zhong, C. Liu, and C. Wu, "Uncovering IT Career Path Patterns with Job Embedding-based Sequence Clustering," *ACM Transactions on Management Information Systems*, vol. 16, iss. 2, pp. 1-32, 2025. **doi: 10.1145/3712705**.

[11] M. Hollander, D.A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*. John Wiley & Sons, Inc., 2013. **doi: 10.1002/9781119196037**.

[12] D.C. Montgomery, E.A. Peck, and G.G. Vining, *Introduction to Linear Regression Analysis*. John Wiley & Sons, Inc., 2013.

[13] A. Bowman, and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford University Press, 1997. **doi: 10.1093/oso/9780198523963.001.0001**.

[14] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts & Techniques*, 3rd ed. Morgan Kaufmann, 2011.

[15] F. Harrel. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, 2001. **doi: 10.1007/978-1-4757-3462-1**.

Стаття надійшла 19.11.2025
Стаття прийнята 02.12.2025
Стаття опублікована 29.12.2025

Цитуйте цю статтю як: Сиротенко О. Г., Петрова Р. В., Морозова А. І. Методи машинного навчання як інструмент аналізу темпів кар'єрного розвитку співробітників ІТ-компанії. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2025. Вип. 52. С. 39-47. DOI: https://doi.org/10.31498/2225-6733.52.2025.350986.