



DOI: 10.58423/2786-6742/2024-6-81-101  
УДК 331:303.732.4

### **Адальберт МАКАРОВИЧ**

здобувач вищої освіти  
за освітньо-науковою програмою «Системи і методи прийняття рішень»  
Київський національний університет імені Тараса Шевченка  
м. Київ, Україна  
**ORCID ID:** 0009-0000-2352-9933

### **Вікторія МАКАРОВИЧ**

кандидат економічних наук, доцент,  
доцент кафедри обліку і аудиту  
Закарпатський угорський інститут імені Ференца Ракоці ІІ,  
м. Берегове, Україна  
**ORCID ID:** 0000-0002-0684-7072  
**ResearchID:** C-1755-2017  
**Scopus Author ID:** 57210844509  
**MTMT:** 10079063

## **МОДЕЛЮВАННЯ ПІДБОРУ МОЛОДИХ СПЕЦІАЛІСТІВ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ**

***Анотація.** Розвиток соціальних мереж, популяризація віддаленої роботи та глобалізація створюють підґрунтя для трансформації HR-менеджменту. Підвищення ефективності та якості реалізації наявних технологій підбору персоналу досягається за рахунок їх цифровізації: застосування інтегрованих мобільних додатків та автоматизація HR-процесів, цифрової інтеграції з хмарними сервісами, використання прогнозної реорле-аналітики, використання технологій доповненої реальності (AR), віртуальної реальності (VR), штучного інтелекту (AI). Однак, наявна проблема пошуку роботи молодими спеціалістами, що полягає у відсутності досвіду, конкуренції та неоднорідності освітніх програм. Через це роботодавцям складно оцінювати та наймати таких кандидатів. Найм молодих спеціалістів важливий для компаній з різних причин, включаючи підготовку майбутніх фахівців та доступ до академічних ресурсів.*

*У статті досліджено потенціал використання методів, що працюють на інтелектуальній основі для підбору молодих спеціалістів, який полягає у автоматизації рутинних завдань, пришвидшення відбору кандидатів, оцінці об'єктивності кандидатури. У рамках проведеного дослідження для прогнозування найму молодих спеціалістів використано методи машинного навчання: логістична регресія; лінійна регресія; дерева рішень; випадковий ліс; XGBoost. Для аналізу даних використано мову програмування R; для тренування та тестування моделей машинного навчання використано мову програмування Python та бібліотеки scikit-learn та XGBoost. При побудові моделей враховано академічну успішність молодих спеціалістів та рівень оплати праці. Результати дослідження можуть бути використані для оптимізації процесу підбору персоналу та рекрутингу молодих спеціалістів для підприємств і організацій. Штучний інтелект може допомогти в автоматизації відбору кандидатів, аналізі їхніх навичок та якіснішій відповідності потребам робочих місць. Дослідження може допомогти молодим спеціалістам краще розуміти вимоги ринку праці, отримати цінні поради та інсайти щодо того, як підготуватися і підвищити власну конкурентоспроможність при пошуку роботи.*



**Ключові слова:** персонал, підбір персоналу, моделювання, штучний інтелект, логістична регресія, лінійна регресія, дерева рішень, випадковий ліс, XGBoost.

**JEL Classification:** M12, O33, O39.

**Absztrakt.** A közösségi hálózatok fejlődése, a távmunka népszerűsítése és a globalizáció megeremti az alapot az HR-menedzsment átalakulásához. A meglévő toborzási technológiák megvalósításának hatékonyságának és minőségének növelése azok digitalizálásával érhető el: integrált mobilalkalmazások alkalmazása és a HR-folyamatok automatizálása, digitális integráció alkalmazása felhőszolgáltatásokkal, prediktív humánelemzés, kiterjesztett valóság (AR), virtuális valóság (VR), ill. mesterséges intelligencia (AI) alkalmazása. Problémát jelent azonban a fiatal szakemberek foglalkoztatása, amely a tapasztalat hiányából, a versenyből és az oktatási programok heterogenitásából áll. Ez megnehezíti a munkáltatók számára az ilyen jelöltek értékelését és felvételét. A fiatal szakemberek vonzása számos okból fontos a vállalatok számára, beleértve a leendő szakemberek képzését és a tudományos forrásokhoz való hozzáférést. A cikk feltárja a fiatal szakemberek kiválasztásának intellektuális alapú módszereinek alkalmazási lehetőségeit, amelyek a rutinfeladatok automatizálásából, a jelöltkiválasztás felgyorsításából és a jelöltség objektivitásának értékeléséből állnak. Az elvégzett kutatás során a következő gépi tanulási módszereket alkalmaztunk a fiatal szakemberek felvételének előrejelzésére: logisztikus regresszió; lineáris regresszió; döntési fák; véletlenszerű erdő; XGBoost. Az R programozási nyelvet használtuk az adatelemzéshez, a Python programozási nyelvet, a scikit-learn és XGBoost könyvtárakat pedig a gépi tanulási modellek betanítására és tesztelésére. A modellek felépítése során a tanulmány figyelembe vette a fiatal szakemberek sikerességi arányát és a javadalmazás szintjét. A kutatás eredményei felhasználhatók a munkaerő-toborzás és a fiatal szakemberek vállalkozásokba, szervezetekbe történő toborzási folyamatának optimalizálására. A mesterséges intelligencia segíthet automatizálni a jelöltek kiválasztását, elemezheti képességeiket, valamint jobban megfelelni a munkakörök igényeinek. A kutatás segíthet a fiatal szakembereknek abban, hogy jobban megértsék a munkaerő-piaci igényeket, értékes tanácsokat kapjanak, és megértsék, hogyan készüljenek fel és növeljék versenyképességüket az álláskeresés során.

**Kulcsszavak:** személyzet, toborzás, modellezés, mesterséges intelligencia, logisztikai regresszió, lineáris regresszió, döntési fák, véletlenszerű erdő, XGBoost.

**Abstract.** The development of social networks, the popularization of remote work, and globalization provide a basis for the transformation of HR management. The efficiency and quality of implementing existing recruitment technologies are enhanced through their digitalization: the use of integrated mobile applications and automation of HR processes, digital integration with cloud services, the use of predictive people analytics, augmented reality (AR), virtual reality (VR), and artificial intelligence (AI). However, there is an existing problem of job search for young professionals, which lies in the lack of experience, competition, and heterogeneity of educational programs. Because of this, employers find it challenging to evaluate and hire such candidates. Hiring young professionals is important for companies for various reasons, including preparing future specialists and accessing academic resources. The article investigates the potential of using intelligent-based methods for recruiting young professionals, which includes automating routine tasks, accelerating candidate selection, and evaluating the objectivity of candidates. The study employs machine-learning methods for predicting the hiring of young professionals: logistic regression, linear regression, decision trees, random forest, and XGBoost. The R programming language was used for data analysis; the Python programming language and the SciKit-Learn and XGBoost libraries were used for training and testing machine-learning models. The models take into account the academic performance of young professionals and their salary levels. The study's results can be used to optimize the recruitment process for enterprises and organizations. Artificial intelligence can assist in automating candidate selection, analyzing their skills, and better matching job requirements. The research can help young professionals



*better understand labor market requirements, gain valuable advice and insights on how to prepare and increase their competitiveness in the job search.*

**Keywords:** *personnel, recruitment, modeling, artificial intelligence, logistic regression, linear regression, decision trees, random forest, XGBoost.*

**Постановка проблеми.** З введенням сучасних технологій в процес підбору персоналу, відкриваються нові можливості для підвищення ефективності та точності відбору нових спеціалістів, а також зниження часових та фінансових витрат на цей процес. Окрім того, у контексті постійних змін на ринку праці та зростаючої конкуренції, дослідження з використання методів, що працюють на інтелектуальній основі для підбору персоналу має важливе значення в рамках розвитку сучасних підходів до управління персоналом та підвищення конкурентоспроможності компаній. Використання новітніх технологій може дати більш неупереджену оцінку кандидатам, передбачити справедливую заробітну плату, а також пришвидшити цей процес. Це зумовлює необхідність поглибленого дослідження підбору персоналу за допомогою статистичних методів та методів, що працюють на інтелектуальній основі.

**Аналіз останніх досліджень та публікацій.** Питання інтеграції нових технологій в HR-менеджмент розглядалися в дослідженнях науковців: Л. Водянка, Д. Ратушняк, О. Лусте, О. Гашков, Р. Золотуха, О. Глазунова, І Іляш, Б. Баб'як, О. Кравчук, А. Кучерков, І. Молнар, С. Кубіцький, Ю. Семенко, О. Скотник, В. Фостолович, Н. Черненко, В. Яценко та ін.

Низка авторів досліджують переваги та потенційні небезпеки застосування штучного інтелекту при підборі персоналу. Зокрема: Н. Черненко акцентує увагу на наступних перевагах застосування штучного інтелекту при підборі персоналу: операційна ефективність, рекрутмент, онбординг, управління талантами, стратегічне планування, кар'єрне зростання й управлінські зміни. Сере ризиків виділяє: неготовність менеджерів до впровадження новітніх технологій в управління людськими ресурсами, етичні проблеми, які можуть виникнути під час використання систем ШІ [6].

Семенко Ю. оцінено переваги та недоліки використання інформаційних технологій та штучного інтелекту в діяльності компанії. Автор зазначає, що штучний інтелект може допомогти ефективно підбирати працівників, оцінювати результати їх роботи, коригувати завдання та призначати навчальні курси на основі аналізу успішності та рівня виконання поставлених перед працівниками завдань. Також використання штучного інтелекту допомагає уникати суб'єктивності в процесі роботи з потенційними працівниками, оскільки штучний інтелект та інформаційні технології дозволяють оцінювати діяльність працівників базуючись, на чітких даних. До недоліків штучного інтелекту, Семенко Ю., в першу чергу, відносить схильність до помилок у базових речах та недосконалість самої технології, що вимагає постійного контролю з боку керівника [3].

Водянка Л., Ратушняк Д., Лусте О. дослідили такі інноваційні методи підбору персоналу, як: хедхантинг (headhunting), цифровий рекрутинг (digital recruitment), скринінг (screening), прямий пошук (executive search), прелімінаринг (preliminaring), аутстафінг (outstaffing), аутплейсмент (outplacement) і лізинг персоналу (personal-leasing). Окреслили характеристику кожного методу, переваги та недоліки використання, а також їх доцільність використання у різних випадках [1]. Інші науковці [2, 5] здійснюють пошук шляхів оптимізації процесу підбору персоналу та ефект від впровадження ІІІ.

Сьогодні ми перебуваємо на перехресті, де важливо визнати перевагу технологічно-орієнтованих процесів над традиційними. Сочинська І., Сторожук О., Доренська А. [4] акцентують, що цифрова трансформація технологій найму працівників прискорює процес руйнування старих бізнес-моделей, спонукає традиційні компанії послідовно переносити застаріле локальне програмне забезпечення в хмарні системи для підтримки сучасних бізнес-процесів і поліпшення якості обслуговування клієнтів.

У свою чергу постачальники корпоративного програмного забезпечення докладають чимало зусиль, щоб підтримати і навіть заохотити до технологічних змін, пропонуючи все більше хмарних пропозицій, причому не лише в галузі ERP (Enterprise Resource Planning – планування ресурсів підприємства), але й хмарне програмне забезпечення для HCM (Human Capital Management - управління людським капіталом) та ТМ (Talent management - управління талантами).

**Метою статті** є дослідження моделювання процесу найму молодих спеціалістів за допомогою статистичних методів та методів, що працюють на інтелектуальній основі.

**Постановка завдання.** Для досягнення поставленої мети визначено наступні завдання: побудова моделей підбору молодих спеціалістів за допомогою методів логістичної регресії, лінійної регресії, дерева рішень, випадкового лісу, XGBoost, їх тестування.

**Викладення основного матеріалу.** Відповідно до дослідження [16] автоматизація та аналітика в процесі рекрутингу та відбору усувають упередження, які зазвичай зустрічаються в ручних методах найму. Як правило, кандидати використовують тактики створення позитивного враження у традиційних інтерв'ю віч-на-віч.

За допомогою автоматизованих процесів рекрутингу можна усунути вплив цих тактик. Штучний інтелект у віртуальних інтерв'ю зменшує людське упередження. Це також допомагає рекрутерам наймати таланти з усього світу.

Однак, дослідники [12], акцентують, що, на сьогодні, недостатньо даних про те, як люди та машини взаємодіють у прийнятті рішень, а також як алгоритми можуть бути ефективно перевірені та регульовані для забезпечення інклюзивності.

Алгоритми машинного навчання можуть значно підсилити процеси найму, пропонуючи швидші, точніші та більш об'єктивні способи оцінки кандидатів. Деякі великі компанії вже почали активно впроваджувати їх. За допомогою ІІІ





можна обробляти інформацію про кандидатів та оцінювати їх, а також навіть шукати потенційних кандидатів, що ще не подали заявки.

Серед існуючих технологій відсутні відомі моделі для оцінювання молодих спеціалістів. Ця сфера є важливою, адже кандидати такого типу дуже мають специфічні характеристики порівняно із іншими претендентами на ринку праці. Наприклад, логічно припустити, що при наймі молодих спеціалістів більш важливою буде їх успішність під час навчання ніж наявний досвід роботи. Розробка моделей, які б полегшили процес найму молодих спеціалістів є перспективним напрямком досліджень.

Для проведення досліджень нами сфокусовано увагу на наймі молодих спеціалістів, які не мають практичного досвіду та взято за основу датасет [8], що містить інформацію про найм студентів в індійській бізнес-школі. Даний датасет зібраний командою відділу розміщення шляхом опитування студентів із університетського містечка після сезону найму, за сприяння доктора Dhimant Ganatara, професора Jain University.

Датасет містить 15 змінних для 215 студентів. В табл. 1 продемонстрована інформація для перших 5 студентів<sup>1</sup>.

**Таблиця 1.**

**Датасет інформації про перших п'ятьох студентів [8]**

sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
1	M	67	Others	91	Others	Commerce	58	Sci&Tech	No	55	Mkt&HR	58,8	Placed	270000
2	M	79,33	Central	78,33	Others	Science	77,48	Sci&Tech	Yes	86,5	Mkt&Fin	66,28	Placed	200000
3	M	65	Central	68	Central	Arts	64	Comm&Mgmt	No	75	Mkt&Fin	57,8	Placed	250000
4	M	56	Central	52	Central	Science	52	Sci&Tech	No	66	Mkt&HR	59,43	Not Placed	
5	M	85,8	Central	73,6	Central	Commerce	73,3	Comm&Mgmt	No	96,8	Mkt&Fin	55,5	Placed	425000

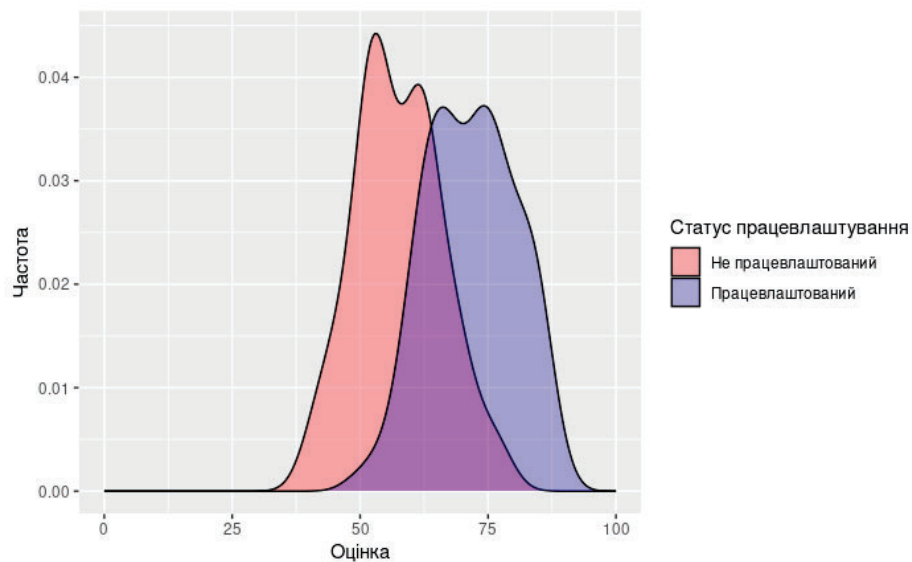
Датасет містить 12 ознак, що характеризують академічну успішність кандидата та тип освіти, який він здобув, а також інформацію про те, чи вдалось кандидату знайти роботу та яку оплату праці він отримав. Поставимо задачу побудувати моделі для передбачення чи здобуде кандидат роботу, та передбачення оплати праці для нього.

Перед початком побудови моделей, важливо проаналізувати значення параметрів та залежності між ними. Такий аналіз допоможе збагнути, що саме

<sup>1</sup>База даних в повному обсязі, яка взята за основу в дослідженні розміщена за посиланням Analysis on Campus Recruitment Data. URL: <https://www.kaggle.com/code/benroshan/you-re-hired-analysis-on-campus-recruitment-data>

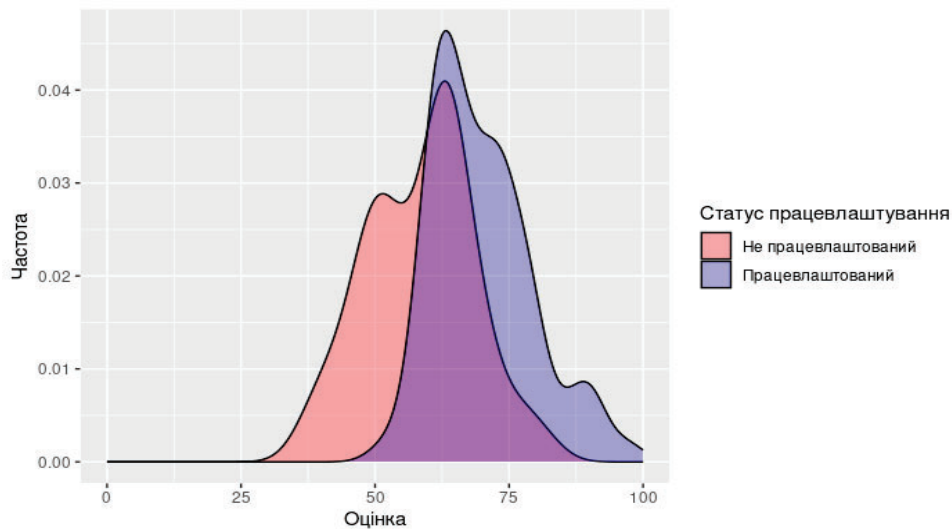
представляють дані, виявити пропущені дані, аномальні значення, виявити які параметри найбільше впливають на цільову функцію і які можуть бути опущеними.

Датасет містить пропущені дані тільки в змінній salary. Заробітна плата не вказана для кандидатів, що не отримали роботу. В датасеті є 67 кандидатів, що так і не були працевлаштовані.



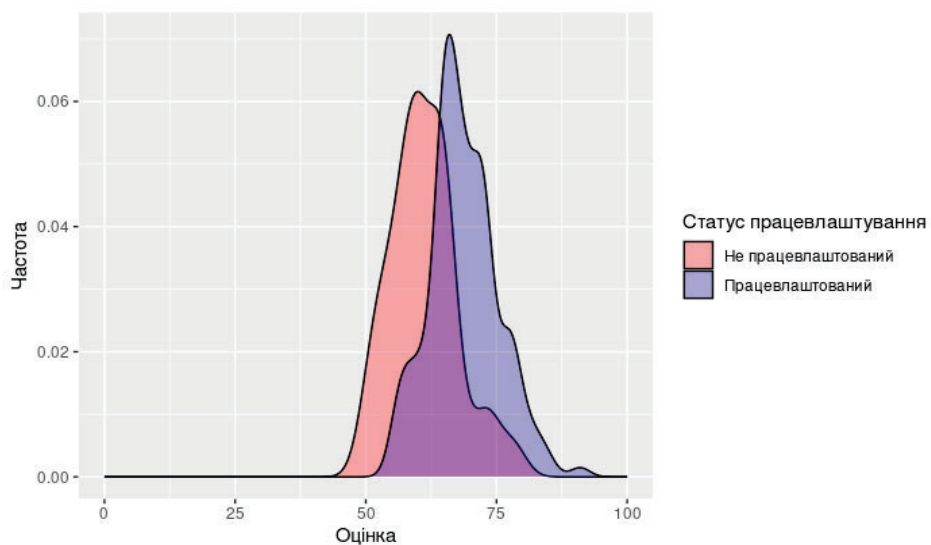
**Рис.1. Розподіл оцінок отриманих особами під час іспитів на атестат середньої освіти**

*Джерело: побудовано авторами за даними [8]*



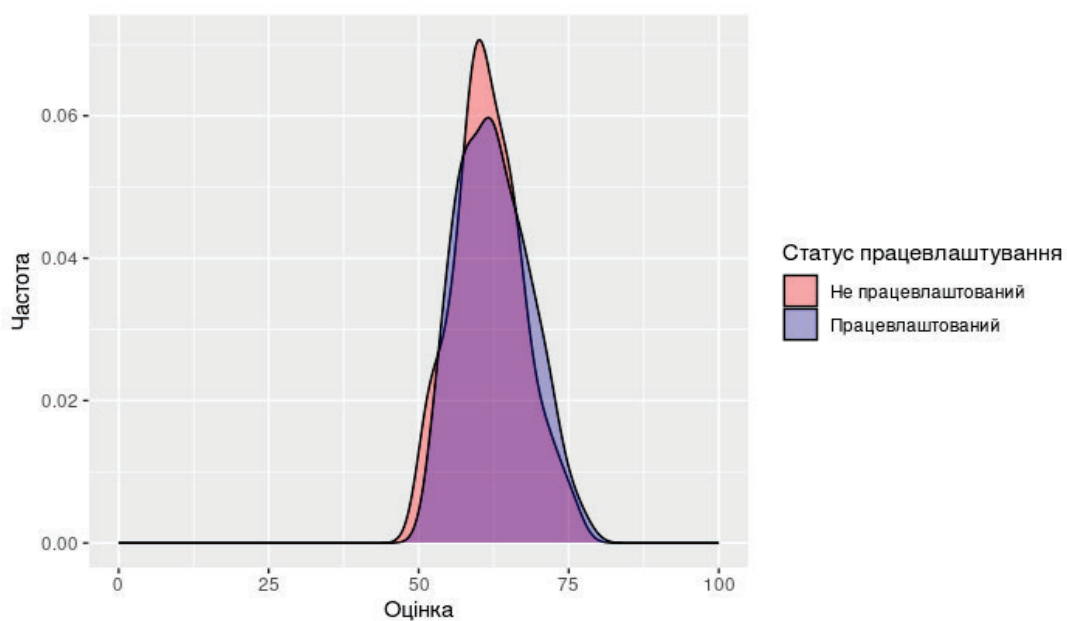
**Рис.2. Розподіл оцінок отриманих особами під час іспитів на атестат вищої середньої освіти**

*Джерело: побудовано авторами за даними [8]*



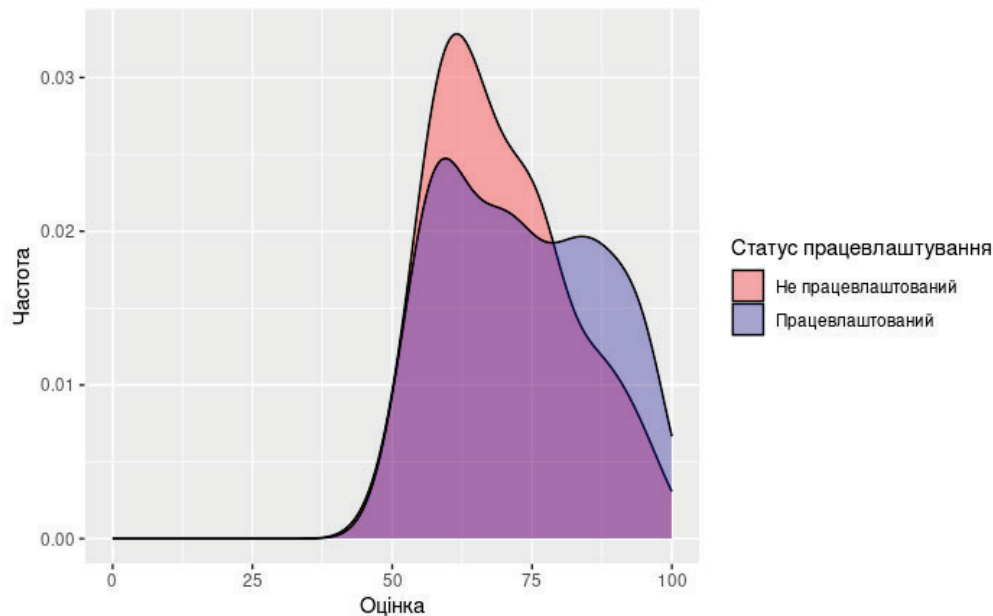
**Рис.3. Розподіл оцінок отриманих особами на першому (бакалаврському) рівні вищої освіти**

*Джерело: побудовано авторами за даними [8]*



**Рис.4. Розподіл оцінок отриманих особами на другому (магістерському) рівні вищої освіти**

*Джерело: побудовано авторами за даними [8]*



**Рис.5. Розподіл оцінок отриманих особами в тесті на працевлаштування**  
Джерело: побудовано авторами за даними [8]

На рис. 1-5 зображено розподіл оцінок кандидатів отриманих під час середньої освіти, вищої середньої освіти, бакалавратури, магістратури та оцінка за тест на працевлаштування для двох категорій кандидатів (ті, хто отримали роботу, та ті які не отримали). Із Рис. 1, 2, 3, 5 можна побачити, що кандидати яким вдалось здобути роботу мають вищі оцінки, отже ці показники можна використовувати для передбачення результату.

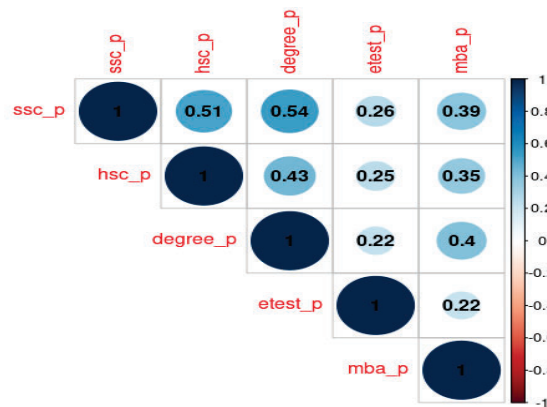
З іншого боку, розподіл оцінок за магістратуру (Рис. 4) майже однаковий для обох груп кандидатів. З цього можна припустити, що ця змінна має слабкий вплив на результат і може бути проігнорована під час тренування моделі.

Серед інформації про успішність кандидата наявні його оцінки на різних етапах навчання (середня освіта, вища середня освіта, бакалаврат, магістратура). Варто перевірити залежність цих показників між собою.

Логічно припустити, що результати навчання на пізніших етапах залежать від базових знань отриманих на ранніх етапах. Кандидат, який добре вивчав математику в школі скоріше за все буде мати хороші успіхи вивчаючи суміжні предмети в університеті.

Побудуємо матрицю парних коефіцієнтів кореляції Пірсона для оцінок. Парна кореляція Пірсона вимірює силу та напрямок лінійних зв'язків між парами безперервних змінних. Кореляція Пірсона оцінює наявність статистичних доказів лінійного зв'язку між однаковими парами змінних у генеральній сукупності [11].





**Рис.6. Матриця парних коефіцієнтів кореляції Пірсона для оцінок**  
 Джерело: розраховано та побудовано авторами за даними [8].

Спираючись на значення парних коефіцієнтів, можна зробити висновок, що існує помірний зв'язок між оцінками за середню освіту, вищу середню освіту та бакалаврат.

В табл. 2 наведені парні коефіцієнти кореляції Пірсона між оцінками та оплатою праці кандидата.

**Таблиця 2.**

**Коефіцієнти кореляції Пірсона між оцінками та оплатою праці кандидата**

Оцінка кандидата за:	Коефіцієнт кореляції Пірсона
атестат середньої освіти	0,14
атестат вищої середньої освіти	0,07
бакалаврат	0,03
магістратуру	0,2
тест на працевлаштування	0,23

Низькі значення коефіцієнта кореляції Пірсона вказують на відсутність лінійного зв'язку між оцінками та оплатою праці. Найбільший коефіцієнт кореляції для оцінки за тест на працевлаштування.

Розглянемо різні моделі машинного навчання для прогнозування того, чи кандидат буде найнятий, та яку оплату праці він отримає. Тестування різних моделей дозволяє порівняти їх та обрати найоптимальнішу для розв'язання поставленої задачі.

*Використання логістичної регресії для передбачення найму кандидатів.*

Методи регресії стали невід'ємною складовою будь-якого аналізу даних, що стосується опису взаємозв'язку між змінною відгуку та однією або кількома пояснювальними змінними. Часто відбувається так, що змінна результату, як у нашому випадку, є дискретною і приймає два значення. Протягом останнього десятиліття модель логістичної регресії стала, у багатьох галузях, стандартним методом аналізу в такій ситуації. Те, що відрізняє модель логістичної регресії від

лінійної моделі регресії, полягає в тому, що змінна результату в логістичній регресії є бінарною або дихотомічною [10].

У логістичній регресії ми моделюємо ймовірність бінарного результату (наприклад, "так" або "ні") в залежності від значень одного чи декількох незалежних змінних. Модель логістичної регресії використовує логістичну функцію (часто називається "сигмоїдною функцією") для трансформації лінійної комбінації вхідних ознак у значення між 0 і 1.

На рис. 7 наведено код для тренування моделі логістичної регресії.

```
# Тренування моделі
logreg = LogisticRegression(max_iter=10000)
logreg.fit(x_train, y_train)
print('Accuracy of logistic regression classifier on train set:
{:.2f}'.format(logreg.score(x_train, y_train)))

# Перевірка результатів моделі
y_pred = logreg.predict(x_test)
confusion_matrix = confusion_matrix(y_test, y_pred)
print('Accuracy of logistic regression classifier on test set:
{:.2f}'.format(logreg.score(x_test, y_test)))
print("Confusion Matrix:\n",confusion_matrix)
print("Classification Report:\n",classification_report(y_test,
y_pred))
```

**Рис. 7. Код для тренування та перевірки результатів логістичної регресії для задачі класифікації**

Дана модель здатна передбачати те, чи буде кандидат найнятий на роботу із точністю 88% як на тренувальному, так і на валідаційному датасеті.

Для тренування використано 10000 ітерацій та алгоритм Limited-memory BFGS [11], який є стандартним для бібліотеки Scikit-learn.

Серед 43 кандидатів валідаційного датасету 2 були помилково класифіковані як не найняті і 3 були помилково класифіковані як найняті.

**Таблиця 3.**

**Матриця невідповідностей для логістичної моделі**

	Не найняті кандидати	Найняті кандидати
Кандидати класифіковані, як не найняті	12	3
Кандидати класифіковані, як найняті	2	26

*Використання лінійної регресії для передбачення оплати праці кандидатів*

Оплата праці є неперервною змінною і має числове значення. Через це, використання логістичної регресії для її передбачення є недоцільним.

Замість логістичної регресії, для передбачення оплати праці використано лінійну регресію.

Основна мета лінійної регресії - знайти лінійну залежність між залежною змінною та однією або декількома незалежними змінними.

На рис. 8 наведено код тренування моделі лінійної регресії.

```
# Тренування моделі
linreg = LinearRegression()
linreg.fit(x_train, y_train)

# Тестування моделі
y_pred = linreg.predict(x_test)

print(f"R^2 score of model is {r2_score(y_test, y_pred)*100} %")
print(f"Root mean squared error is {sqrt(mean_squared_error(y_test,
y_pred))}")
print(f"Mean absolute error is {mean_absolute_error(y_test, y_pred)}")
```

**Рис. 8. Код для побудови та перевірки результатів лінійної регресії для задачі передбачення оплати праці**

Як показники точності моделі вибрано наступні параметри:

а) коефіцієнт детермінації - міра відповідності моделей, яка базується на частці поясненої дисперсії [7].

б) корінь середньоквадратичної похибки

в) середня абсолютна похибка

В табл. 4 наведено значення цих показників для моделі лінійної регресії.

**Таблиця 4.**

**Значення показників точності прогнозування зарплатні для лінійної регресії**

Коефіцієнт	Значення
коефіцієнт детермінації	87,34%
корінь середньоквадратичної похибки	53494,53
середня абсолютна похибка	33749,24

*Використання дерева рішень для передбачення найму кандидатів та оплати праці*

Методологія дерева рішень - це широко використовуваний метод в галузі дата-майнінгу для створення класифікаційних систем на основі кількох варіантів.

Цей метод класифікує вибірку на гілкові сегменти, які формують інвертоване дерево з кореневим вузлом, внутрішніми вузлами та листками.

Тренувальний набір даних використовується для побудови моделі дерева рішень. Валідаційний набір даних використовується для визначення оптимального розміру дерева, необхідного для досягнення оптимальної кінцевої моделі та для перевірки результатів моделі [13].

На рис. 9. наведено код для побудови та тестування моделі для передбачення статусу найму кандидатів.

```
# Тренування моделі
dectree = DecisionTreeClassifier(criterion = 'gini', max_depth=3)
dectree = dectree.fit(x_train, y_train)
print('Accuracy of decision tree classifier on train set:
{:.2f}'.format(dectree.score(x_train, y_train)))

# Перевірка результатів моделі
y_pred = dectree.predict(x_test)
confusion_matrix = confusion_matrix(y_test, y_pred)
print('Accuracy of decision tree classifier on test set:
{:.2f}'.format(dectree.score(x_test, y_test)))
print("Confusion Matrix:\n",confusion_matrix)
```

**Рис. 9. Код для побудови та перевірки результатів дерева рішень для задачі класифікації**

Модель показує найкращі результати при глибині дерева - 3.

Як функцію вимірювання якості розбиття використано gini. Дана модель здатна передбачати те, чи буде кандидат найнятий на роботу із точністю 90% на тренувальному датасеті та 74% на валідаційному датасеті.

Серед 43 кандидатів валідаційного датасету 2 були помилково класифіковані як не найняті і 9 були помилково класифіковані як найняті.

Такий класифікатор призводить до ризиків, що кандидати без необхідних кваліфікацій будуть найняті на роботу.

Саме такий тип помилок вигідно уникати роботодавцю.

**Таблиця 5.**

**Матриця невідповідностей для дерева рішень**

	Не найняті кандидати	Найняті кандидати
Кандидати класифіковані, як не найняті	6	9
Кандидати класифіковані, як найняті	2	26

На основі ілюстрації дерева, наведеного на рис. 10. можна побачити, що класифікація відбувається на основі оцінок за середню освіту, вищу середню освіту та диплом бакалавра.

Хоча дана модель не дає гірший результат, але вона легка в інтерпретації. Інструкції дерева рішень можуть бути використані рекрутерами під час відбору кандидатів та проведення співбесід.

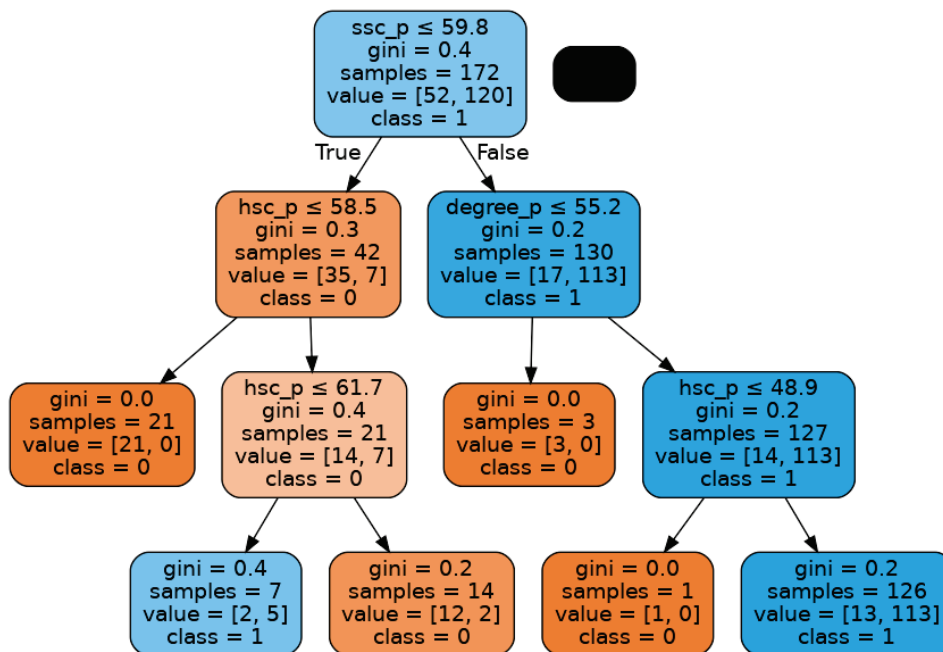


Рис.10. Дерево рішень для класифікації кандидатів за статусом працевлаштування

На рис. 11. наведено код для побудови та тестування моделі для передбачення оплати праці кандидатів.

```
# Тренування моделі
dectree = DecisionTreeRegressor()
dectree.fit(x_train, y_train)

# Тестування моделі
y_pred = dectree.predict(x_test)

print(f"R^2 score of model is {r2_score(y_test, y_pred)*100} %")
print(f"Root mean squared error is {sqrt(mean_squared_error(y_test, y_pred))}")
print(f"Mean absolute error is {mean_absolute_error(y_test, y_pred)}")
```

Рис. 11. Код для побудови та перевірки результатів дерева рішень для задачі передбачення оплати праці

Результати використання дерева рішень для прогнозування оплати праці наведені в табл. 6.



Таблиця 6.

**Значення показників точності прогнозування зарплатні для дерева рішень**

Коефіцієнт	Значення
коефіцієнт детермінації	50,64%
корінь середньоквадратичної похибки	105651.66
середня абсолютна похибка	43454.55

*Використання випадкового лісу для передбачення найму кандидатів*

Оскільки, одне дерево не може створити достатньо точне передбачення, спадає на думку ідея спробувати використати випадковий ліс. Агрегація результатів декількох дерев може дати більш точну модель.

Випадковий ліс — це комбінація дерев рішень, у яких кожне дерево залежить від значень випадкового вектора, відібраного незалежно та з однаковим розподілом для всіх дерев у лісі [9].

На рис. 12. наведено код для тренування випадкового лісу. Дана модель дає 100% точність на тренувальному датасеті, але лише 79% на валідаційному.

```
# Тренування моделі
ranfor = RandomForestClassifier(n_estimators=10)
ranfor = ranfor.fit(x_train, y_train)
print('Accuracy of random forrest classifier on train set:
{:.2f}'.format(ranfor.score(x_train, y_train)))

# Перевірка результатів моделі
y_pred = ranfor.predict(x_test)
confusion_matrix = confusion_matrix(y_test, y_pred)
print('Accuracy of random forrest classifier on test set:
{:.2f}'.format(ranfor.score(x_test, y_test)))
print("Confusion Matrix:\n",confusion_matrix)
```

**Рис. 12. Код для побудови та перевірки результатів випадкового лісу для задачі класифікації**

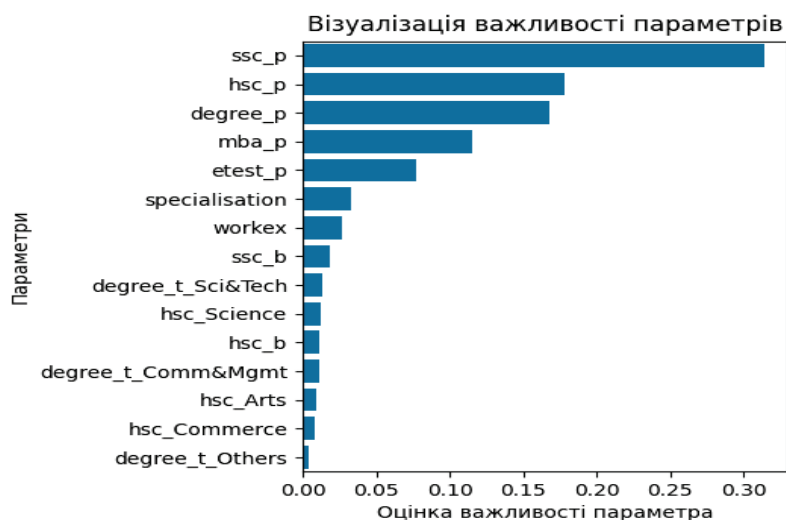
Результати використання методу випадкового лісу для наведені в табл. 7.

Таблиця 7.

**Матриця невідповідностей для випадкового лісу**

	Не найняті кандидати	Найняті кандидати
Кандидати класифіковані, як не найняті	7	8
Кандидати класифіковані, як найняті	1	27

На рис. 13 візуалізовано рівень важливості вхідних параметрів. Відкинувши найменш важливі з них можна спробувати уникнути проблеми перенавчання.



**Рис13. Візуалізація рівня важливості параметрів у випадковому лісі**

Доволі дивно, що оцінка за середню школу є найголовнішим параметром. Після видалення інформації про галузь ступеня бакалавра, спеціалізацію вищої середньої освіти та освітні ради модель почала показувати трохи кращі результати (табл. 8.). Точність моделі - 81%.

**Таблиця 8.**

**Матриця невідповідностей для випадкового лісу після видалення деяких параметрів**

	Не найняті кандидати	Найняті кандидати
Кандидати класифіковані, як не найняті	8	7
Кандидати класифіковані, як найняті	1	27

На рис. 14 наведено код для побудови та тестування моделі для передбачення зарплатні кандидатів.

```
# Тренування моделі
ranfor = RandomForestRegressor()
ranfor.fit(x_train, y_train)

# Тестування моделі
y_pred = ranfor.predict(x_test)

print(f"R^2 score of model is {r2_score(y_test, y_pred)*100} %")
print(f"Root mean squared error is {sqrt(mean_squared_error(y_test, y_pred))}")
print(f"Mean absolute error is {mean_absolute_error(y_test, y_pred)}")
```

**Рис. 14. Код для побудови та перевірки результатів випадкового лісу для задачі передбачення оплати праці**

Результати використання випадкового лісу для прогнозування зарплатні наведені в табл. 9.

Таблиця 9.

**Значення показників точності прогнозування зарплатні для випадкового лісу**

Коефіцієнт	Значення
коефіцієнт детермінації	93,80%
корінь середньоквадратичної похибки	37455.26
середня абсолютна похибка	21740.45

*Використання XGBoost для передбачення найму кандидатів*

Підсилення дерев - це високоефективний і широко використовуваний метод машинного навчання. XGBoost - масштабована система підсилення дерев, яка широко використовується аналізі та багатьох задачах машинного навчання. Цей алгоритм, враховує розрідженість даних та ваговий квантильний зріз для наближеного навчання дерев. Комбінуючи інформацію про шаблони доступу до кешу, стиснення даних та розподілу, XGBoost масштабується, використовуючи набагато менше ресурсів, ніж альтернативні системи [15].

На рис. 15. наведено код для тренування та тестування XGBoost для прогнозування статусу найму.

```
# Тренування моделі
xgreg = xgb.XGBClassifier(objective = 'reg:logistic', learning_rate =
0.1, n_estimators = 100, alpha = 1, max_depth = 5)
xgreg.fit(x_train,y_train)
y_pred = xgreg.predict(x_train)
print('Accuracy of XGBoost classifier on train set:
{:.2f}'.format(accuracy_score(y_train, y_pred) * 100))

# Перевірка результатів моделі
y_pred = xgreg.predict(x_test)
print('Accuracy of XGBoost classifier on train set:
{:.2f}'.format(accuracy_score(y_test, y_pred) * 100))
y_pred = pd.Series(y_pred, index = y_test.index)
confusion_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n",confusion_matrix)
```

**Рис. 15. Код для побудови та перевірки результатів XGBoost для задачі класифікації**

Модель натренована за допомогою XGBoost дає точність 99.42% при прогнозуванні статусу найму на тренувальних даних та 90.70% на валідаційних. Серед 43 кандидатів валідаційного датасету 1 були помилково класифіковані як не найняті і 3 були помилково класифіковані як найняті (табл. 10).



Таблиця 10.

**Матриця невідповідностей для XGBoost**

	Не найняті кандидати	Найняті кандидати
Кандидати класифіковані, як не найняті	12	3
Кандидати класифіковані, як найняті	1	27

На рис. 16. наведено код для тренування та тестування XGBoost для прогнозування зарплатні.

```
# Тренування моделі
xgreg = xgb.XGBRegressor()
xgreg.fit(x_train, y_train)

# Тестування моделі
y_pred = xgreg.predict(x_test)

print(f"R^2 score of model is {r2_score(y_test, y_pred)*100} %")
print(f"Root mean squared error is {sqrt(mean_squared_error(y_test, y_pred))}")
print(f"Mean absolute error is {mean_absolute_error(y_test, y_pred)}")
```

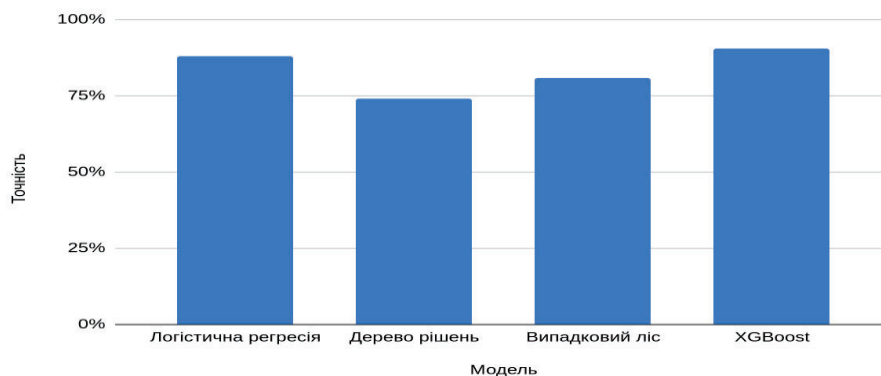
**Рис. 16. Код для побудови та перевірки результатів XGBoost для задачі передбачення оплати праці**

Результати використання XGBoost для прогнозування зарплатні наведені в табл. 11.

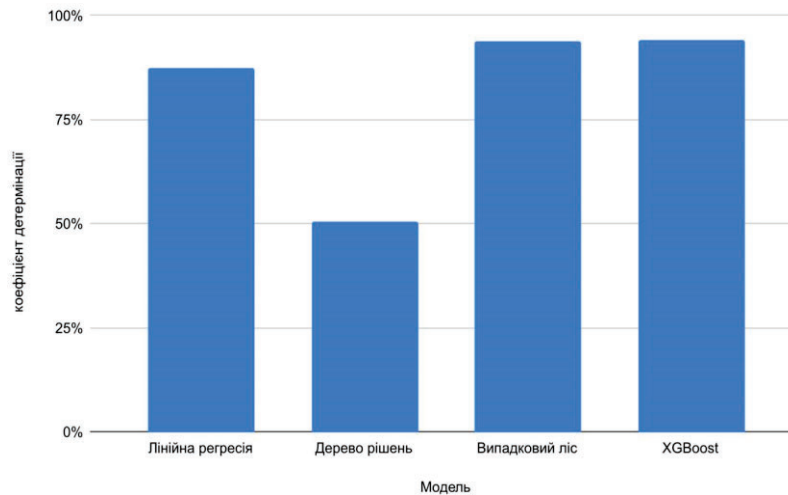
Таблиця 11.

**Значення показників точності прогнозування зарплатні для XGBoost**

Коефіцієнт	Значення
коефіцієнт детермінації	94,09%
корінь середньоквадратичної похибки	36551.07
середня абсолютна похибка	21433.74



**Рис.17. Гістограма точності передбачення статусу кандидатів**



**Рис.18. Гістограма коефіцієнта детермінації передбачення оплати праці кандидатів**

Як можна побачити із рис. 17,18 найкращу точність для обох задач дає XGBoost.

Моделі на основі XGBoost передбачають статус працевлаштування кандидата із 90% точністю, а середня абсолютна похибка передбачення оплати праці складає 21433.74 грошових одиниць в рік. Така висока точність прогнозування дає надію на те, що модель можна використовувати як допоміжний спосіб для прийняття рішень про найм кандидатів. Наприклад, з її допомогою можна проводити попередню фільтрацію кандидатів. Варто зауважити, що оскільки моделі натреновані на датасеті індійських студентів, які працюють у сфері ділового адміністрування, їх використання в інших умовах є обмеженим.

**Висновки та пропозиції.** Моделювання підбору молодих спеціалістів за допомогою новітніх технологій дає можливість ефективніше підібрати та використовувати людські ресурси. Аналізуючи структуру та зв'язки між даними, можна зрозуміти важливі аспекти функціонування системи та забезпечити її ефективну роботу.

Для побудови моделей нами розглянуто датасет зібраний командою відділу розміщення шляхом опитування студентів із університетського містечка після сезону найму. Датасет містить інформацію про академічну успішність 215 кандидатів та їх статус працевлаштування. Проаналізовано дані наявні в датасеті та залежності між ними.

В рамках дослідження використано наступні методи машинного навчання для прогнозування статусу працевлаштування кандидата та його оплати праці: логістична регресія, лінійна регресія, дерева рішень, випадковий ліс, XGBoost. Найкращі результати отримано при використанні XGBoost.

За результатами дослідження можна зробити висновок, що використання методів машинного навчання, зокрема XGBoost, для передбачення найму





молодих спеціалістів на основі їх академічної успішності має значний потенціал у вдосконаленні процесів набору персоналу. Цей підхід дозволяє автоматизувати рутинні завдання, зменшує час на відбір кандидатів та може сприяти об'єктивнішому відбору. Зважаючи на успішність використання штучного інтелекту в компаніях, таких як Unilever та Facebook, для оптимізації процесів найму та розширення діапазону кандидатів, можна очікувати, що подальший розвиток цих технологій принесе значні переваги як для роботодавців, так і для кандидатів, сприяючи підвищенню ефективності та об'єктивності найму молодих спеціалістів.

Попри великий потенціал, застосування штучного інтелекту для найму молодих спеціалістів, дані технології мають використовуватись як допоміжні і строго контролюватись людиною, щоб уникнути небезпек пов'язаних із дискримінацією певних груп кандидатів та контролювати загальну коректність алгоритмів. Важливим є регулярність перегляду та оновлення моделей ШІ для рекрутингу, задля забезпечення їх об'єктивності та точності.

Подальші дослідження можуть бути сфокусовані на дослідженні використання аналогічних моделей машинного навчання із датасетами із інших країн та спеціальностей.

### Список використаних джерел

1. Водянка, Л., Ратушняк, Д., & Лусте, О. (2022) Інноваційні методи підбору персоналу в умовах диджиталізації. *БІЗНЕСІНФОРМ № 1 '2022*, 403-409. URL: <https://doi.org/10.32983/2222-4459-2022-1-403-409> (дата звернення 23.03.2024).
2. Золотуха, Р., & Глазунова, О. (2023). Проектування схеми бази даних для процесу автоматизації підбору персоналу в ІТ команди. *Collection of Scientific Papers «SCIENTIA»*, (November 3, 2023; Bern, Switzerland), 127–129. URL: <https://previous.scientia.report/index.php/archive/article/view/1294> (дата звернення 21.03.2024).
3. Семененко Ю. (2024) Роль інформаційних технологій та інструментів штучного інтелекту в підвищенні ефективності підбору, навчання та адаптації працівників. *Галицький економічний вісник*, № 2 (87). 20-29. URL: [https://doi.org/10.33108/galicianvisnyk\\_tntu2024.02](https://doi.org/10.33108/galicianvisnyk_tntu2024.02) (дата звернення 10.04.2024).
4. Сочинська-Сибірцева, І., Сибірцева, О., & Доренська, А. Новітні технології управління персоналом: навч. посіб. Кропивницький : ЦНТУ, 2023. 278 с.
5. Фостолович В., Боцян Т., Павлова С., Фостолович Р., Гуртовий О. (2023) Штучний інтелект у сфері гостинності: місце інтегрування, специфіка використання та вплив на доходи підприємства. *Економіка. Управління. Інновації* Випуск №1 (32). URL: <http://eprints.zu.edu.ua/37678/1/283100-Article%20Text-652510-1-10-20230627.pdf> (дата звернення 23.03.2024).
6. Черненко, Н. (2022). Штучний інтелект в управлінні персоналом. *Таврійський науковий вісник. Серія: Економіка*, (12), 76-83. URL: <https://doi.org/10.32851/2708-0366/2022.12.11> (дата звернення 23.03.2024).
7. Alessandro Di Bucchianico. Coefficient of Determination (R2). *Wiley StatsRef: Statistics Reference Online*. URL: <https://doi.org/10.1002/9780470061572.eqr173> (дата звернення 04.04.2024).
8. Analysis on Campus Recruitment Data. URL: <https://www.kaggle.com/code/benroshan/you-rehired-analysis-on-campus-recruitment-data> (дата звернення 10.04.2024).



9. Breiman L. (2001) Random Forests. *Machine Learning*. Volume 45, pp.5–32. URL: <https://doi.org/10.1023/A:1010933404324> (дата звернення 20.04.2024).
10. David W. Hosmer, Jr., Stanley Lemeshow. Applied Logistic Regression. URL: [https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/REGRESI%20LOGISTIK/epdf.pub\\_applied-logistic-regression-wiley-series-in-probab.pdf](https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/REGRESI%20LOGISTIK/epdf.pub_applied-logistic-regression-wiley-series-in-probab.pdf) (дата звернення 10.03.2024).
11. Liu D. C. & Nocedal J. (1989) On the Limited Memory Method for Large Scale Optimization. *Mathematical Programming B*. 45 (3). 503–528. URL: 10.1007/BF01589116. S2CID 5681609
12. Pawan Budhwar, Soumyadeb Chowdhury, Geoffrey Wood, Herman Aguinis, Greg J. Bamber, Jose R. Beltran, Paul Boselie, Fang Lee Cooke, Stephanie Decker, Angelo DeNisi, Prasanta Kumar Dey, David Guest, Andrew J. Knoblich, Ashish Malik, Jaap Paauwe, Savvas Papagiannidis, Charmi Patel, Vijay Pereira, Shuang Ren, Steven Rogelberg, Mark N. K. Saunders, Rosalie L. Tung, Arup Varma. (2023) Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT, *Human Resource Management Journal*. 33, 3. (606-659). URL: <https://doi.org/10.1111/1748-8583.12511>
13. Song YY, & Lu Y. (2015) Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*. Apr 25 27(2);130-5. URL: 10.11919/j.issn.1002-0829.215044.
14. SPSS TUTORIALS: PEARSON CORRELATION. URL: <https://libguides.library.kent.edu/SPSS/PearsonCorr> (дата звернення 10.04.2024).
15. Tianqi Chen, Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. URL: <https://dl.acm.org/doi/pdf/10.1145/2939672.2939785> (дата звернення 10.04.2024).
16. Tyagi, P., Chilamkurti, N., Grima, S., Sood, K. and Balusamy, B. (Ed.) (2023) The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part A, *Emerald Publishing Limited*, Leeds. pp. i-xxii. URL: <https://doi.org/10.1108/978-1-80382-027-920231015> (дата звернення 10.04.2024).

## References

1. Vodianka, L., Ratuszniak, D., & Luste, O. (2022) Innovatsiini metody pidboru personalu v umovakh dydzhyalizatsii. *BIZNESINFORM* № 1 '2022, 403-409. URL: <https://doi.org/10.32983/2222-4459-2022-1-403-409> (last accessed: 23.03.2024).
2. Zolotukha, R., & Hlazunova, O. (2023). Proektuvannia skhemy bazy danykh dlia protsesu avtomatyzatsii pidboru personalu v IT komandy. *Collection of Scientific Papers «SCIENTIA»*, (November 3, 2023; Bern, Switzerland), 127–129. URL: <https://previous.scientia.report/index.php/archive/article/view/1294> (last accessed: 21.03.2024).
3. Semenenko Yu. (2024) Rol informatsiinykh tekhnolohii ta instrumentiv shtuchnoho intelektu v pidvyshchenni efektyvnosti pidboru, navchannia ta adaptatsii pratsivnykiv. *Halytskyi ekonomichnyi visnyk*, № 2 (87). 20-29. URL: [https://doi.org/10.33108/galicianvisnyk\\_tntu2024.02](https://doi.org/10.33108/galicianvisnyk_tntu2024.02) (last accessed: 10.04.2024).
4. Sochynska-Sybirtseva, I., Sybirtseva, O., & Dorenska, A. Novitni tekhnolohii upravlinnia personalom: navch. posib. Kropyvnytskyi : TsNTU, 2023. 278 s.
5. Fostolovych V., Botsian T., Pavlova S., Fostolovych R., Hurtovyi O. (2023) Shtuchnyi intelekt u sferi hostynnosti: mistse intehruvannia, spetsyfika vykorystannia ta vplyv na dokhody pidpriemstva. *Ekonomika. Upravlinnia. Innovatsii Vypusk* №1 (32). URL: <http://eprints.zu.edu.ua/37678/1/283100-Article%20Text-652510-1-10-20230627.pdf> (last accessed: 23.03.2024).
6. Chernenko, N. (2022). Shtuchnyi intelekt v upravlinni personalom. *Tavriiskyi naukovyi visnyk. Serii: Ekonomika*, (12), 76-83. URL: <https://doi.org/10.32851/2708-0366/2022.12.11> (last accessed: 23.03.2024).
7. Alessandro Di Bucchianico. Coefficient of Determination (R2). *Wiley StatsRef: Statistics Reference Online*. URL: <https://doi.org/10.1002/9780470061572.eqr173> (дата звернення 04.04.2024).



8. Analysis on Campus Recruitment Data. URL: <https://www.kaggle.com/code/benroshan/you-re-hired-analysis-on-campus-recruitment-data> (дата звернення 10.04.2024).
9. Breiman L. (2001) Random Forests. *Machine Learning*. Volume 45, pp.5–32. URL: <https://doi.org/10.1023/A:1010933404324> (last accessed: 20.04.2024).
10. David W. Hosmer, Jr., Stanley Lemeshow. Applied Logistic Regression. URL: [https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/REGRESI%20LOGISTIK/epdf.pub\\_applied-logistic-regression-wiley-series-in-probab.pdf](https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/REGRESI%20LOGISTIK/epdf.pub_applied-logistic-regression-wiley-series-in-probab.pdf) (last accessed: 10.03.2024).
11. Liu D. C. & Nocedal J. (1989) On the Limited Memory Method for Large Scale Optimization. *Mathematical Programming B*. 45 (3). 503–528. URL: 10.1007/BF01589116. S2CID 5681609
12. Pawan Budhwar, Soumyadeb Chowdhury, Geoffrey Wood, Herman Aguinis, Greg J. Bamber, Jose R. Beltran, Paul Boselie, Fang Lee Cooke, Stephanie Decker, Angelo DeNisi, Prasanta Kumar Dey, David Guest, Andrew J. Knoblich, Ashish Malik, Jaap Paauwe, Savvas Papagiannidis, Charmi Patel, Vijay Pereira, Shuang Ren, Steven Rogelberg, Mark N. K. Saunders, Rosalie L. Tung, Arup Varma. (2023) Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT, *Human Resource Management Journal*. 33, 3. (606-659). URL: <https://doi.org/10.1111/1748-8583.12511>
13. Song YY, & Lu Y. (2015) Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*. Apr 25 27(2);130-5. URL: 10.11919/j.issn.1002-0829.215044.
14. SPSS TUTORIALS: PEARSON CORRELATION. URL: <https://libguides.library.kent.edu/SPSS/PearsonCorr> (last accessed: 10.04.2024).
15. Tianqi Chen, Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. URL: <https://dl.acm.org/doi/pdf/10.1145/2939672.2939785> (last accessed: 10.04.2024).
16. Tyagi, P., Chilamkurti, N., Grima, S., Sood, K. and Balusamy, B. (Ed.) (2023) The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part A, *Emerald Publishing Limited*, Leeds. pp. i-xxii. URL: <https://doi.org/10.1108/978-1-80382-027-920231015> (last accessed: 10.04.2024).