КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет інформаційних технологій

Кафедра технологій управління

Спеціальність 122 – Комп'ютерні науки, освітня програма «Інформаційна аналітика та впливи»

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему:

«РОЗРОБКА ТЕХНОЛОГІЇ УПРАВЛІННЯ ПЕРСОНАЛОМ МЕТОДАМИ DATA SCIENCE»

Студентки 2-го курсу групи ІАВ-21

Сніжани Михайлівни КНИШ
(прізвище, ім'я, по батькові)

_____ _____ _____
(підпис студента) (дата) (підпис)

Науковий керівник:

д.т.н., доц.
(науковий ступінь, вчене звання)

Юлія Леонідівна ХЛЕВНА
(прізвище, ім'я, по батькові)

Попередній захист:		
(Висновок: «До захисту в Екзаменаційній комісії»)		
Завідувач кафедри технологій управління	Віктор МОРОЗОВ	
(підпис)	(прізвище, ініціали)	(дата)

Київ – 2025

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет інформаційних технологій

Кафедра технологій управління
Освітньо-кваліфікаційний рівень Магістр
Спеціальність 122 – Комп'ютерні науки
Освітня програма Інформаційна аналітика та впливи

ЗАТВЕРДЖУЮ
Завідувач кафедри, професор
Морозов В.В.
_____
«_____» 20___ року

ЗАВДАННЯ НА ВИКОНАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Студент Сніжана КНИШ

Група ІАВ-21

1. Тема кваліфікаційної роботи

Розробка технології управління персоналом методами Data Science

Затверджено наказом від «___» _____ 20___ р. № _____

2. Строк подання студентом готової роботи – "___" _____ 20___ р.

3. Цільова установка та вихідні дані до роботи

Розробити прогнозну модель для оцінки термінів закриття вакансій із використанням методів машинного навчання у середовищі R, орієнтовану на підвищення точності планування HR-процесів. Основою дослідження є набір структурованих рекрутингових даних, що містять характеристики вакансій, кандидата та джерел пошуку. Передбачено використання алгоритмів ансамблевого навчання (зокрема Random Forest), обґрунтування вибору метрик оцінки моделі, аналіз впливу змінних та візуалізація результатів.

4. Зміст роботи

Аналіз сучасних підходів до застосування методів Data Science у сфері управління персоналом, зокрема у прогнозуванні термінів закриття вакансій.

Формалізація задачі регресійного моделювання на основі рекрутингових даних та обґрунтування вибору відповідного алгоритму машинного навчання (Random Forest).

Підготовка та обробка вхідних даних, включаючи категоріальні змінні, перевірка якості даних і усунення ознак витоку інформації.

Побудова, навчання та тестування прогнозної моделі.

Оцінка точності моделі з використанням метрик MAE та RMSE, аналіз впливу факторів на прогноз.

Інтерпретація результатів моделювання у контексті HR-аналітики та оцінка практичної цінності отриманих висновків.

5. Перелік графічного матеріалу (слайдів)

Робота містить 1 аналітичну таблицю та 23 рисунки, серед яких – графіки розподілу змінних, ілюстрації результатів обробки рекрутингових даних, діаграми важливості ознак, графік розсіювання прогнозованих і фактичних значень, порівняння показників MAE та RMSE, а також візуальні схеми реалізації моделі прогнозування у HR-процесах.

6. Календарний план виконання роботи

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва частин роботи	%	Виконання роботи	
			За планом	Фактично
1.	Вибір теми дипломної роботи	3	01.10.24	01.10.24
2.	Протокол кафедри ТУ про затвердження тем дипломних робіт та призначення наукових керівників	2	27.12.24	27.12.24
3.	Формування переліку нормативних матеріалів, літератури з проблематики дипломної роботи	10	08.01.25	07.01.25
4.	Складання розгорнутого плану кваліфікаційної роботи	5	18.01.25	18.01.25
5.	Ознайомлення наукового керівника з розгорнутим планом кваліфікаційної роботи. Внесення змін.	5	19.01.25 - 20.01.25	20.01.25
6.	Підготовка розділу 1 «Data Science у сфері управління персоналом»	10	12.02.25	13.02.25
7.	Підготовка розділу 2 «Формалізація технології прогнозування термінів закриття вакансій із використанням технологій data science»	14	08.03.25	08.03.25

8.	Підготовка розділу 3 «Розробка методу data science для прогнозування термінів закриття вакансій»	14	20.03.25	20.03.25
9.	Підготовка розділу 4 «Технологія впровадження моделі прогнозування термінів закриття вакансій у практичну діяльність»	13	15.04.25	15.04.25
10.	Оформлення кваліфікаційної роботи. Підготовка висновків і пропозицій	15	25.04.25	25.04.25
11.	Передача кваліфікаційної роботи науковому керівникові	2	01.05.25	01.05.25
12.	Передача кваліфікаційної роботи рецензенту для рецензування	2	04.05.25	04.05.25
13.	Попередній захист кваліфікаційної роботи	5	13.05.25	13.05.25

Дата видачі завдання «___» _____ 2025 р.

Керівник роботи д.т.н., доц. Юлія Леонідівна ХЛЕВНА _____

(посада, прізвище, ім’я, по батькові) (підпис)

Завдання прийняла до виконання студентка групи ІАВ-21: Сніжана КНИШ

_____

(підпис)

ЗМІСТ

АНОТАЦІЯ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет інформаційних технологій

Кафедра технологій управління

Спеціальність 122 - Комп'ютерні науки,

освітня програма "Інформаційна аналітика та впливи"

Дипломна робота магістра Сніжани КНИШ.

Тема роботи – «Розробка технології управління персоналом методами Data Science».

Мета дипломної роботи магістра – підвищення ефективності рекрутингових процесів у системі управління персоналом шляхом розробки та впровадження технології прогнозування термінів закриття вакансій із застосуванням методів Data Science.

Об'єкт дослідження – процеси Data Science в сфері управління персоналом, спрямовані на прогнозування термінів закриття вакансій та оптимізацію рекрутингових процесів.

Предмет дослідження – моделі, методи та технології Data Science, технології управління персоналом методами Data Science.

Наукова новизна дослідження полягає полягає у розширенні предметного поля HR-аналітики шляхом формалізації підходу до прогнозування тривалості закриття вакансій як самостійного об'єкта дослідження, що раніше не набув належного наукового висвітлення та розглядався здебільшого в контексті загальної оцінки ефективності рекрутингу. У роботі обґрунтовано доцільність виокремлення цього аспекту як окремої задачі моделювання в межах застосування методів Data Science. Відмінність запропонованого підходу полягає в орієнтації на прогнозування саме термінів закриття вакансій, що залишається малодослідженим аспектом у сучасній HR-аналітиці, на відміну від широко представлених рішень щодо класифікації кандидатів або оцінки їхньої

відповідності. Модель враховує поєднання категоріальних та кількісних змінних, специфічних для рекрутингового процесу, а також адаптована для застосування у внутрішніх HR-системах без потреби у складних технічних інтеграціях.

Практична цінність цього дослідження полягає у розробці та впровадженні технології прогнозування термінів закриття вакансій на основі методів Data Science. Запропонована технологія дозволяє компаніям оптимізувати процеси рекрутингу, скорочуючи час пошуку кандидатів та ефективніше розподіляючи HR-ресурси. Результати дослідження можуть бути використані для вдосконалення систем планування кадрових потреб, що сприяє своєчасному укомплектуванню команд проєктів і зменшенню ризиків затримок у виконанні бізнес-процесів. Крім того, розроблена технологія може слугувати основою для створення аналітичних HR-дашбордів, які забезпечують прозорість і прогнозованість процесів найму. Застосування моделі також дає змогу підвищити якість управлінських рішень у сфері людських ресурсів, забезпечуючи організаціям гнучкість та адаптивність до змін на ринку праці.

Дипломна робота складається зі вступу, основної частини, яка включає чотири розділи, висновків та переліку використаних джерел. Всього налічує 95 сторінок, перелік посилань з 37 джерел на 5 сторінках.

Ключові слова: управління персоналом, Data Science, рекрутинг, Random Forest, прогнозування, HR-аналітика, машинне навчання, аналітична модель.

ПЕРЕЛІК ВИКОРИСТАНИХ СКОРОЧЕНЬ

HR – Human Resources / управління персоналом

HRM – Human Resource Management / управління людськими ресурсами

HRIS – Human Resource Information System / інформаційна система управління персоналом

DSS – Decision Support System / система підтримки прийняття рішень

MAE – Mean Absolute Error / середня абсолютна похибка

RMSE – Root Mean Square Error / корінь з середньої квадратичної похибки

IT – Information Technology / інформаційні технології

R – мова програмування R

GDPR – General Data Protection Regulation / Загальний регламент захисту даних (ЄС)

GNN – Graph Neural Networks / графові нейронні мережі

ВСТУП

Управління персоналом (HRM) є ключовим елементом ефективного функціонування будь-якої організації. В умовах сучасного бізнес-середовища, яке характеризується високою конкуренцією та швидкими змінами, стратегічне управління людськими ресурсами набуває все більшого значення. Успішне управління персоналом сприяє досягненню організаційних цілей, підвищенню продуктивності та зниженню витрат. Однак, традиційні методи управління персоналом часто не можуть задовольнити потреби сучасних організацій, що викликає необхідність впровадження нових технологій і підходів.

Наука про дані (Data Science) є однією з таких новітніх технологій, що пропонує значні переваги для управління персоналом. Data Science – це міждисциплінарна галузь, яка використовує наукові методи, алгоритми та системи для вилучення знань і інсайтів з даних у різних формах, як структурованих, так і неструктурованих. Застосування Data Science у сфері управління персоналом відкриває нові можливості для прийняття обґрунтованих рішень, оптимізації процесів та підвищення ефективності організації.

Впровадження Data Science у HRM дозволяє аналізувати великі обсяги даних про співробітників, що допомагає виявляти приховані закономірності та тенденції. Це, в свою чергу, сприяє покращенню процесу відбору та найму, управлінню талантами, аналізу продуктивності та залученості, прогнозуванню плинності кадрів та розробці стратегій компенсації і пільг. За допомогою аналітичних інструментів можна більш точно визначати потреби організації, оптимізувати витрати на персонал та створювати умови для залучення і утримання висококваліфікованих працівників, що і зумовлює **актуальність** цієї кваліфікаційної **роботи.**

Таким чином, наше дослідження націлене на поглиблене розуміння того, як наука про дані може трансформувати управління персоналом, сприяючи досягненню стратегічних цілей організації та підвищенню її конкурентоспроможності.

Зв'язок роботи з науковими програмами, планами, темами: кваліфікаційна робота виконана в межах реалізації освітньо-наукової програми «Інформаційна аналітика та впливи», що передбачає формування компетенцій у сфері прикладного аналізу даних, машинного навчання та підтримки управлінських рішень. Тематика роботи узгоджується з цілями навчальної програми та базується на результатах, отриманих під час науково-дослідної практики у HR-відділі IT-компанії, де досліджувались актуальні проблеми оптимізації рекрутингових процесів засобами Data Science. Робота також пов'язана з виконанням науково-дослідної теми кафедри: «Розробка інформаційно-аналітичних інструментів управління портфелями проєктів і програм в інтегрованих функціональних середовищах» (державний реєстраційний номер 0121U107799), яка реалізується у Київському національному університеті імені Тараса Шевченка.

Мета кваліфікаційної роботи: підвищення ефективності рекрутингових процесів у системі управління персоналом шляхом розробки та впровадження технології прогнозування термінів закриття вакансій із застосуванням методів Data Science.

Завдання кваліфікаційної роботи:

- Проаналізувати теоретичні та практичні аспекти використання Data Science в управлінні персоналом.
- Визначити та обрати моделі і методи Data Science для вирішення визначених проблем.
- Реалізувати обрані моделі та методи на мові програмування R.
- Провести аналіз отриманих результатів та оцінити ефективність використаних моделей у HR-процесах.

Об'єкт дослідження: процеси Data Science в сфері управління персоналом, спрямовані на прогнозування термінів закриття вакансій та оптимізацію рекрутингових процесів.

Предметом дослідження кваліфікаційної роботи є моделі, методи та технології Data Science, технології управління персоналом методами Data Science.

Для досягнення мети дослідження застосовано сукупність **методів**, що забезпечують як теоретичне обґрунтування, так і практичну реалізацію підходу до прогнозування термінів закриття вакансій. Теоретичну базу було сформовано за допомогою аналізу наукових джерел, що стосуються застосування Data Science у сфері управління персоналом, із метою виявлення сучасних тенденцій, інструментів та наукових прогалин. Для формалізації задачі прогнозування використано методи системного аналізу та концептуального моделювання, що дозволило структурувати ключові чинники, які впливають на тривалість рекрутингового процесу.

На етапі побудови моделі використано методи машинного навчання, зокрема алгоритм Random Forest, який було реалізовано засобами мови програмування R. Для попередньої обробки даних застосовано методи факторизації категоріальних змінних, масштабування, крос-валідації та стратифікованого розбиття вибірки. Оцінювання якості моделі здійснювалося за допомогою показників MAE та RMSE. Інтерпретацію результатів проведено із використанням методів оцінки важливості змінних, що забезпечило виявлення ключових факторів, пов’язаних із термінами закриття вакансій.

Наукова новизна дослідження полягає полягає у розширенні предметного поля HR-аналітики шляхом формалізації підходу до прогнозування тривалості закриття вакансій як самостійного об’єкта дослідження, що раніше не набув належного наукового висвітлення та розглядався здебільшого в контексті загальної оцінки ефективності рекрутингу. У роботі обґрунтовано доцільність виокремлення цього аспекту як окремої задачі моделювання в межах застосування методів Data Science. Відмінність запропонованого підходу полягає в орієнтації на прогнозування саме термінів закриття вакансій, що залишається малодослідженим аспектом у сучасній HR-аналітиці, на відміну від широко представлених рішень щодо класифікації кандидатів або оцінки їхньої

відповідності. Модель враховує поєднання категоріальних та кількісних змінних, специфічних для рекрутингового процесу, а також адаптована для застосування у внутрішніх HR-системах без потреби у складних технічних інтеграціях.

Практична цінність цього дослідження полягає у розробці та впровадженні технології прогнозування термінів закриття вакансій на основі методів Data Science. Запропонована модель дозволяє компаніям оптимізувати процеси рекрутингу, скорочуючи час пошуку кандидатів та ефективніше розподіляючи HR-ресурси. Результати дослідження можуть бути використані для вдосконалення систем планування кадрових потреб, що сприяє своєчасному укомплектуванню команд проєктів і зменшенню ризиків затримок у виконанні бізнес-процесів. Крім того, розроблена технологія може слугувати основою для створення аналітичних HR-дашбордів, які забезпечують прозорість і прогнозованість процесів найму. Застосування моделі також дає змогу підвищити якість управлінських рішень у сфері людських ресурсів, забезпечуючи організаціям гнучкість та адаптивність до змін на ринку праці.

Результати роботи апробовано на Міжнародній науковій конференції *Information Technology and Implementation (Satellite): Conference Proceedings*, 20-21 листопада 2023 року, м. Київ; за підсумками представлено одну публікацію. Ключові положення дипломного дослідження також стали основою наукової роботи, що була відзначена як переможець першого туру Всеукраїнського конкурсу студентських наукових робіт з галузей знань і спеціальностей у 2024/2025 навчальному році.

РОЗДІЛ 1
DATA SCIENCE У СФЕРІ УПРАВЛІННЯ ПЕРСОНАЛОМ

1.1 Концепція науки про дані та її актуальність у сфері управління персоналом

В умовах розвитку інформаційного суспільства в різних секторах економіки постійно створюється та накопичується значний обсяг різноманітних даних. У промисловості та бізнесі безперервно збільшується потік інформації, необхідної для ефективного управління підприємствами. Постійно з'являються нові сервіси, що базуються на застосуванні інформаційних та комунікаційних технологій. Внаслідок розвитку інтернету, соціальних мереж, відео-, аудіо- та геолокаційних сервісів безперервно зростає попит на інформаційні продукти та послуги.

Для надання таких послуг підприємствам доводиться аналізувати великі обсяги даних із різноманітних джерел. Для державних органів, телекомунікаційних та інтернет-компаній, банків, підприємств роздрібної торгівлі, енергетичного сектору та житлово-комунального господарства накопичена інформація стає стратегічно важливим активом. Ефективне управління цим активом суттєво впливає на результати їхньої діяльності. Зростання обсягів інформації супроводжується розвитком апаратних і програмних засобів, здатних оперативно обробляти великі обсяги даних, а також значним зниженням вартості збору, обробки, зберігання та передачі інформації.

У результаті з'єднання цих двох процесів – зростання потреби бізнесу в обробці та зберіганні великих обсягів даних і появи технічних засобів, здатних оперативно обробляти такі дані з мінімальними витратами – виник один із найцікавіших і перспективних напрямів розвитку послуг, відомий як Big Data (великі дані) [36].

Data Science, або наука про дані, є міждисциплінарною галуззю, яка об'єднує методи статистики, математики, інформатики та машинного навчання

для аналізу великих обсягів даних [35]. Основні компоненти Data Science включають збір, обробку та аналіз даних, машинне навчання, візуалізацію даних і прогнозну аналітику.

У сфері управління персоналом (HR) Data Science відкриває нові можливості для покращення ефективності HR-процесів та прийняття обґрунтованих рішень. Аналіз даних дозволяє здійснювати глибоке вивчення різноманітних аспектів роботи співробітників, від їхньої продуктивності до рівня задоволеності. Завдяки Data Science, HR-фахівці можуть виявити внутрішні та зовнішні фактори, що впливають на робочий процес, та розробляти стратегії для їх оптимізації. Інтеграція аналітичних інструментів дозволяє не лише виявляти проблеми в управлінні персоналом, а й швидко реагувати на них, направляючи ресурси та зусилля в найбільш ефективні напрямки [35]. Таким чином, Data Science стає не просто інструментом аналізу даних, але й ключовим компонентом у формуванні стратегічних рішень в галузі управління персоналом. Розвиток науки про дані та машинного навчання запровадив нові перспективні методи розуміння й аналізу структури компанії та динаміки робочої сили. Завдяки аналітиці організації можуть покращити спосіб пошуку, найму та утримання співробітників, що робить аналіз кадрових даних вирішальним для бізнес-діяльності.

Сфера управління персоналом переходить в епоху наукового аналізу, що базується на передових технологіях, відходячи від традиційних методів, таких як прості опитування та психологічні оцінки. Замість того, щоб покладатися на нескінченні опитування чи пряме спілкування зі співробітниками для оцінки лояльності та задоволеності роботою, цінні контрольні показники можна отримати завдяки методам великих даних [10]. Спеціалісти, які працюють з даними, підтверджують актуальність, важливість і надійність корпоративних даних в сфері управління персоналом. Вони допомагають фахівцям з управління персоналом структурувати свою аналітику, усвідомлюючи, що деякі застарілі показники можуть неточно відображати такі фактори, як утримання співробітників або довгострокову задоволеність місцем роботи [1]. Експерти з

обробки даних можуть відстежувати та впорядковувати зібрані дані, аналізувати відповідні набори даних і представляти свої висновки за допомогою зрозумілих графіків і діаграм, які чітко передають актуальні дані і можливі наслідки.

Очевидно, що сьогоднішній процес найму суттєво трансформується і стає більш складним завдяки новим технологіям і підходам. Сучасні компанії замість традиційного підходу, що включав відбір кандидатів за резюме та співбесідами, активно використовують великі дані для аналізу масивів інформації з різних джерел. Це дозволяє їм створювати інформативні профілі кандидатів, оцінювати їхні навички та прогнозувати їхній успіх у компанії з набагато більшою точністю [13]. Крім того, наука про дані дозволяє HR-фахівцям не лише залучати кращих кандидатів, але й оптимізувати всі аспекти процесу найму та управління персоналом загалом. Аналізуючи дані щодо ефективності рекрутингових кампаній, витрат на навчання нових працівників та плинність кадрів, компанії можуть ефективніше управляти ресурсами та зменшувати витрати [21]. Завдяки аналізу великих обсягів даних, фахівці можуть глибше розуміти потреби своєї компанії та ефективно відстежувати ключові показники успішності і, відповідно, оперативно впроваджувати необхідні зміни.

Як показує дослідження Deloitte, 90% HR-фахівців налаштовані на трансформацію своїх організаційних моделей, оновлюючи структури управління та розвитку кадрів [35]. Використання науки про дані стає ключовим фактором у цих процесах, допомагаючи вдосконалювати стратегії залучення талантів, підвищувати ефективність навчальних програм, аналізувати явище плинності кадрів та адаптувати процеси найму до поточних потреб компанії.

Заміна застарілих методів аналізу метрик у сфері управління людськими ресурсами за допомогою науки про дані значно просуває галузь HR, надаючи компаніям глибокі інсайти, які традиційні опитування та інтерв'ю не здатні забезпечити. Таким чином, наука про дані перетворює традиційний HR-підхід у стратегічний інструмент, що сприяє не лише залученню, а й утриманню талановитих співробітників, що є критично важливим для успіху будь-якої сучасної організації.

1.2 Аналіз джерел щодо застосування Data Science у сфері управління персоналом

Застосування методів Data Science у сфері управління персоналом є одним із ключових напрямів сучасних досліджень, що спрямовані на оптимізацію процесів найму, утримання персоналу та підвищення ефективності управлінських рішень. Використання аналітики великих даних дозволяє компаніям оцінювати ефективність рекрутингових стратегій, прогнозувати поведінку кандидатів і співробітників, а також автоматизувати ухвалення рішень у сфері HR.

У цьому розділі буде проведено аналіз існуючих наукових досліджень і літературних джерел, що стосуються застосування технологій Data Science у сфері управління персоналом (HR) та рекрутингу. Data Science набуває все більшої актуальності у сучасних організаціях, оскільки дозволяє обробляти великі обсяги даних для прийняття обґрунтованих рішень, підвищення ефективності процесів і оптимізації управління людськими ресурсами. Особлива увага буде приділена питанням прогнозування термінів закриття вакансій, оцінки ефективності різних каналів пошуку кандидатів, а також аналізу існуючих методів і технік аналізу даних у цій сфері. Метою аналізу є виявлення існуючих підходів, оцінка їхніх переваг і недоліків, визначення досягнутих результатів та перспектив розвитку, а також виявлення прогалин у дослідженнях, що обґрунтовують необхідність подальшого вивчення цієї проблематики. Для аналізу даних у сфері управління персоналом (HR) використовуються різноманітні методи та техніки Data Science. Ось деякі з них:

1. Описова аналітика: зосереджується на створенні звітів, які підсумовують діяльність компанії: статичні знімки бізнес-активності та транзакцій, що надаються особам, які приймають рішення, на регулярній основі (щоденно, щотижнево, щоквартально); постійні перегляди бізнес-ефективності, зазвичай представлених графічно у вигляді інформаційних панелей [13]. Методи, що використовуються в рамках описової аналітики:

Статистичний аналіз: Застосування базових статистичних методів (середнє значення, медіана, мода, дисперсія, стандартне відхилення) для опису основних характеристик даних про персонал, таких як середня зарплата, відсоток працівників різних статей, середній вік тощо [20]. Статистичний аналіз дозволяє отримати базове уявлення про стан справ у компанії, зрозуміти розподіл даних і виявити основні тенденції та відхилення.

Візуалізація даних: Використання графіків, діаграм і табличних звітів для наочного представлення інформації, наприклад, для побудови гістограм для аналізу розподілу зарплат, створення лінійних графіків для відстеження змін чисельності працівників з плином часу, використання діаграм кругових і стовпчастих для порівняння відсотків працівників різних категорій Візуалізація даних дозволяє швидко зрозуміти основні тренди, взаємозв'язки та аномалії в даних. Інструменти візуалізації, такі як Tableau, Power BI [26] надають можливість створювати інтерактивні та динамічні звіти, що полегшують аналіз і прийняття рішень.

2. Діагностична аналітика: зосереджується на розумінні причин і факторів, що впливають на різні аспекти бізнесу, зокрема на управління персоналом. Вона дозволяє глибше аналізувати дані та виявляти взаємозв'язки між різними змінними [7]. Основними методами діагностичної аналітики є кореляційний аналіз та регресійний аналіз.

Кореляційний аналіз: Використовується для визначення, чи існує зв'язок між різними аспектами діяльності працівників, наприклад, між рівнем задоволеності роботою та продуктивністю. Він допомагає зрозуміти, які фактори можуть бути взаємопов'язані, що може бути корисним для подальшого аналізу і прийняття рішень.

Регресійний аналіз: Використовується для оцінки впливу різних факторів на певний результат, наприклад, як освіта, досвід та робоче навантаження впливають на продуктивність працівників. Регресійний аналіз допомагає не тільки виявити наявність зв'язків, але й оцінити їх силу і значущість, що дає можливість розробити цільові заходи для покращення показників [13].

3. Прогнозна аналітика: зосереджується на використанні історичних даних для передбачення майбутніх подій та тенденцій. У сфері управління персоналом це дозволяє компаніям робити обґрунтовані прогнози щодо поведінки працівників, потреб у персоналі та інших важливих показників [17]. Основними методами прогнозної аналітики є машинне навчання та прогнозування на основі моделей часового ряду

Машинне навчання: Алгоритми машинного навчання, такі як регресія, класифікація, дерева рішень, використовуються для прогнозування майбутніх подій, наприклад, відтоку працівників або ймовірності успішності кандидатів. Машинне навчання дозволяє виявляти складні і нелінійні взаємозв'язки у великих обсягах даних, що забезпечує точніші та надійніші прогнози.

Прогнозування: Використання моделей часового ряду для передбачення майбутніх потреб у персоналі, змін у рівні зарплат та інших метрик. Моделі часового ряду дозволяють враховувати сезонні коливання, тренди та інші тимчасові закономірності, що підвищує точність прогнозів [23].

4. Прескриптивна аналітика: зосереджується на розробці конкретних рекомендацій для дій на основі даних і моделей. У сфері управління персоналом це дозволяє компаніям не лише прогнозувати майбутні події, але й визначати найкращі способи реагування на ці події [12]. Основними методами прескриптивної аналітики є оптимізаційні моделі та сценарне моделювання.

Оптимізаційні моделі: Використання лінійного програмування та інших оптимізаційних методів для розробки рекомендацій щодо оптимального розподілу ресурсів, планування графіків роботи та управління талантом. Лінійне програмування дозволяє визначити найкращі способи використання обмежених ресурсів для досягнення максимальних результатів, наприклад, мінімізувати витрати на персонал або максимізувати продуктивність.

Сценарне моделювання: метод аналізу, який передбачає створення різних можливих сценаріїв розвитку подій та оцінку їх впливу на організацію. В сфері управління персоналом може використовуватись для оцінки впливу можливих

рішень на організацію, розробки стратегій та прийняття обґрунтованих рішень. Сценарне моделювання дозволяє компаніям передбачати наслідки різних дій та вибирати найкращий варіант [23]. Наприклад, аналіз різних стратегій найму персоналу або впровадження нових політик може допомогти вибрати найбільш ефективний підхід.

5. Текстова аналітика зосереджується на аналізі текстових даних для виявлення корисної інформації. У сфері управління персоналом це дозволяє аналізувати відгуки співробітників, опитування задоволеності та інші текстові джерела, щоб зрозуміти настрої та проблеми, з якими стикаються працівники [30]. Основними методами текстової аналітики є аналіз відгуків співробітників та сентимент-аналіз.

Аналіз відгуків співробітників: Використання методів обробки природної мови (NLP) для аналізу текстових даних, таких як відгуки співробітників, опитування задоволеності та соціальні мережі. Дозволяє виявити основні теми, проблеми та настрої у відгуках співробітників, що може допомогти у виявленні проблем та прийнятті відповідних рішень.

Сентимент-аналіз: метод, що дозволяє виявляти настрої та емоції співробітників щодо певних аспектів роботи або корпоративної культури. Сентимент-аналіз дозволяє швидко отримати уявлення про загальний настрій у компанії, виявити потенційні проблеми або позитивні аспекти, що потребують уваги [24].

Сучасні дослідження підтверджують, що використання методів аналітики даних та штучного інтелекту у сфері HR значно покращує ефективність процесу рекрутингу, підвищуючи точність відбору кандидатів та оптимізуючи використання ресурсів. Алгоритми машинного навчання все частіше застосовуються для аналізу резюме, оцінки компетенцій кандидатів та прогнозування ключових HR-метрик, що дозволяє компаніям ухвалювати більш обґрунтовані рішення. Так, застосування Data Science у сфері рекрутингу позитивно впливає на ефективність процесу найму. Зокрема, результати регресійного аналізу демонструють, що це значно підвищує загальну

продуктивність рекрутингових процесів ($\beta = 0.58$, $p < 0.001$) та покращує взаємодію з кандидатами ($\beta = 0.61$, $p < 0.001$) завдяки автоматизованим механізмам відбору [27]. Наприклад, деякі науковці [19] пропонують чотирьохступеневу аналітичну модель із застосуванням методу Random Forest для автоматизованого аналізу резюме кандидатів та прогнозування їх відповідності вакансії, що дозволяє скоротити час рекрутингу та підвищити його ефективність.

Інше дослідження [2] розширює застосування предиктивної аналітики у HR та окрім Random Forest застосовує і Gradient Boosting, що дозволяє підвищити точність прогнозування відповідності кандидатів вакансіям, враховуючи ширший набір змінних. Таким чином, вчені фокусуються на виявленні ключових факторів, що впливають на успішність кандидата та, відповідно, прогнозуванні ефективності потенційних співробітників.

Цікавий підхід використано в роботі Рахіма та Квана [29], де запропоновано модель прогнозування заробітної плати на основі ключових навичок кандидатів та інших змінних, застосовуючи методи регресійного аналізу та дерева рішень, що дозволяє оцінювати вплив технічних і поведінкових компетенцій на конкурентоспроможність кандидатів у сфері Data Science.

Фахівці продовжують розширювати межі застосування Data Science у сфері HR, зосереджуючи увагу на визначенні найефективніших методів аналізу кандидатів та прийняття рішень у процесі рекрутингу. У цьому контексті Фраццетто та ін. досліджують можливості використання графових нейронних мереж (GNNs) для виявлення прихованих патернів у великих наборах даних кандидатів, що дозволяє не лише оцінювати їхні навички, але й визначати потенційну відповідність до корпоративної культури та команди. Аналогічно, вчені аналізують, як анкетні опитування можуть стати альтернативою традиційному аналізу резюме, демонструючи, що відповідність кандидата очікуванням роботодавця значно краще прогнозується через оцінку поведінкових характеристик, ніж через структурований аналіз CV [9]. Такий підхід поєднує методи предиктивної аналітики та когнітивного моделювання,

допомагаючи рекрутерам робити не лише статистично обґрунтовані прогнози, а й приймати більш адаптивні рішення, що враховують широкий спектр змінних

Якщо більшість досліджень використовують обмежений набір даних, то робота Пессааха та ін. базується на унікальному великому наборі даних, що містить сотні тисяч рекрутингових записів за понад 10 років та зосереджена на оптимізації процесу рекрутингу шляхом поєднання предиктивної аналітики та математичного програмування. Основна мета дослідження – розробка аналітичної системи підтримки прийняття рішень для рекрутингу, яка не лише прогнозує ймовірність успіху кандидатів, а й оптимізує вибір кандидатів на основі бізнес-обмежень компанії [28]. Автори застосовують моделі байєсівської мережі змінного порядку для аналізу історичних даних про кандидатів.

Розглядаючи аспекти HR-аналітики, значна кількість досліджень присвячена вивченню можливостей використання аналітики даних для підтримки ухвалення стратегічних рішень, зокрема через впровадження прогнозних моделей та HR-дешбордів. У цьому контексті Хамієддін та ін. розглядають роль прогнозної аналітики у підвищенні адаптивності HR-управління, зосереджуючись на автоматизації збору та аналізу ключових HR-метрик, що дозволяє компаніям оперативно реагувати на зміни у сфері найму, утримання персоналу та розвитку компетенцій [15].

Продовжуючи цю тенденцію, Тулі та ін. окреслюють ключові можливості предиктивної аналітики як інструменту для підвищення ефективності HR-управління. Зокрема, вона дозволяє прогнозувати рівень плинності кадрів, що дає змогу компаніям завчасно розробляти стратегії утримання персоналу та мінімізувати ризики втрати цінних співробітників. Крім того, предиктивна аналітика сприяє оптимізації процесу найму, забезпечуючи оцінку відповідності кандидатів до вакансій на основі історичних даних та поведінкових патернів. Окрему увагу автор приділяє можливості аналізу продуктивності співробітників, що допомагає ідентифікувати високопотенційних фахівців та ефективніше керувати розвитком талантів у компанії [33]. Також прогнозні моделі сприяють

адаптації HR-стратегій до змін ринку праці, дозволяючи організаціям формувати більш гнучку кадрову політику.

Додатково, деякі вчені досліджують застосування статистичних методів для аналізу залежностей між HR-метриками та ефективністю процесів управління персоналом [22]. Використання лінійних регресійних моделей дозволяє визначити, які фактори найбільше впливають на швидкість закриття вакансій, рівень залученості працівників та їхню довготривалу продуктивність. Автори підкреслюють, що багатофакторний аналіз, який включає такі змінні, як тип вакансії, рівень заробітної плати, джерела пошуку кандидатів та попередній досвід найму, дозволяє HR-відділам робити більш точні прогнози та адаптувати кадрові стратегії відповідно до змін ринку праці.

Таким чином, аналіз літературних джерел засвідчує активний розвиток застосування методів Data Science у сфері управління персоналом, особливо для оптимізації рекрутингових процесів, оцінки ефективності каналів найму та прогнозування плинності кадрів. Водночас встановлено, що питання прогнозування термінів закриття вакансій, яке має критичне значення для планування ресурсів і забезпечення безперервності бізнес-процесів, досліджується недостатньо. Більшість наявних робіт зосереджені на загальній оцінці ефективності рекрутингу або підборі кандидатів, тоді як розробка моделей для точного прогнозування часу найму залишається актуальним і перспективним напрямом подальших досліджень.

1.3 Збір та управління даними. Джерела даних релевантних для сфери управління персоналом

У сучасному динамічному бізнес-середовищі роль збору та управління даними у сфері управління людськими ресурсами (HR) все більше стає ключовою для успіху організацій. Здатність збирати, аналізувати та використовувати дані перетворила практики управління персоналом від традиційного до стратегічного процесу прийняття рішень. Цей розділ досліджує

фундаментальне значення збору та управління даними у сфері HR, підкреслюючи його роль у прийнятті обґрунтованих рішень та оптимізації стратегій управління працівниками. Збір даних у сфері HR охоплює широкий спектр процесів, що включає систематичне отримання інформації про співробітників, метрики організаційного успіху та зовнішні ринкові дані. Шляхом використання даних з різних джерел, таких як інформаційні системи управління людськими ресурсами (HRIS), платформи для рекрутингу, опитування співробітників та звіти, відділи управління кадрами отримують цінну інформацію про залучення співробітників, найм талантів та ефективність організації. У цьому розділі ми детально розглянемо джерела даних, релевантних для сфери управління персоналом та методи їх збору.

В аналітиці управління людськими ресурсами (HR) джерела даних є ключовими для організацій, щоб мати змогу збирати та аналізувати відповідну інформацію, яка дозволить приймати обґрунтовані рішення. Можемо виокремити такі загальноприйняті джерела даних, релевантні для сфери управління людськими ресурсами:

Внутрішні джерела даних:

- Інформаційні системи управління людськими ресурсами (HRIS): HRIS функціонує як певна база даних, що містить інформацію про співробітників, таку як персональні дані, досвід роботи, оцінки продуктивності, деталі щодо оплати тощо.
- Опитування: опитування серед працівників, такі як анкетування залученості, оцінка задоволеності або пульсові опитування, збирають відгуки про досвід співробітників, їхнє сприйняття та включеність у внутрішню структуру організації [1]. Аналіз результатів опитувань дає змогу краще зрозуміти настрої працівників і виявити області, де можливі поліпшення.
- Системи обліку робочого часу: дані про відвідуваність працівників, записи відпусток та робочі години. Ця інформація дає змогу відслідковувати, як працівники використовують свій робочий час, виявити певні тенденції.

- Дані про заробітну плату: інформація про виплати працівникам, включаючи зарплати, бонуси і соціальні вигоди, що важливо для фінансового управління і компенсаційної політики.
- Оцінки продуктивності: такі оцінювання роботи працівників або команди проводяться з метою визначення їхньої ефективності, досягнень цілей та виявлення можливостей для покращень [21].

Зовнішні джерела даних:

- Соціальні медіа: платформи соціальних мереж можуть стати цінним джерелом даних для аналітики управління людськими ресурсами. Моніторинг соціальних медіа дозволяє організаціям отримувати уявлення про бренд роботодавця, настрій співробітників і виявлення нових тенденцій у сфері управління талантами.
- Портали вакансій та рекрутингові платформи: надають дані про поточні вакансії, профілі кандидатів і тенденції в рекрутингу, допомагаючи при пошуку та залученні талантів.
- Державні бази даних: містять статистику ринку праці, регулювання та демографічні дані, що допомагають у плануванні та відповідності законодавству.
- Ринкові тенденції: аналіз ринкових звітів, тенденцій та економічних показників допомагають організаціям зрозуміти фактори, які впливають на залучення талантів, утримання співробітників та загальні стратегії управління людськими ресурсами [37].

Зі зростанням значення аналітики даних у сфері HR, організації отримують можливість більш ефективно прогнозувати потреби в персоналі та розробляти ефективні стратегії найму, утримувати та управляти персоналом. Доступ до різних джерел інформації, від внутрішніх HRIS до зовнішніх соціальних мереж та державних баз даних, є вирішальним для цього. Загальний успіх в імплементації стратегій кадрового управління часто залежить від комплексного підходу до збору, аналізу та використання даних, що відображають різноманітні аспекти роботи колективу та його взаємодію з організацією.

1.4 Стратегії ефективного збору даних: питання конфіденційності та забезпечення якості даних

У сфері управління персоналом питання конфіденційності даних є надзвичайно важливим. HR-відділи обробляють великі обсяги персональних даних співробітників, і недотримання законодавчих вимог щодо їх захисту може мати серйозні наслідки, як юридичні, так і репутаційні.

В Україні захист персональних даних регулюється Законом України «Про захист персональних даних», прийнятим у 2010 році. Цей закон визначає правові та організаційні основи захисту персональних даних і має на меті захист прав і свобод людини. Основні положення закону включають вимогу отримання згоди суб'єкта персональних даних на їх обробку, забезпечення захисту даних від несанкціонованого доступу та обов'язок власників даних вжити всіх необхідних заходів для забезпечення їх безпеки [37]. У Європейському Союзі захист персональних даних регулюється Загальним регламентом про захист даних (GDPR), який набув чинності у 2018 році. GDPR є одним із найсуворіших законодавчих актів у сфері захисту даних у світі. Основні вимоги GDPR включають отримання явної згоди суб'єкта даних на їх обробку, право суб'єктів на доступ до своїх даних та їх видалення (право на забуття), а також вимогу інформувати суб'єктів про порушення безпеки даних у найкоротший термін. Організації також зобов'язані призначити відповідального за захист даних (DPO), якщо обробка даних є основною діяльністю компанії. Порушення вимог GDPR може призвести до значних штрафів, що можуть досягати до 20 мільйонів євро або 4% від річного глобального обороту компанії [6]. Крім України та ЄС, інші країни також мають свої законодавчі вимоги щодо захисту персональних даних. Наприклад, у Сполучених Штатах Америки існують різні закони на федеральному рівні та рівні штатів, зокрема Каліфорнійський закон про захист прав споживачів (CCPA) [5]. Ці закони мають свої вимоги до обробки та захисту персональних даних, надаючи суб'єктам даних право контролювати використання їхньої інформації.

У випадку ІТ-компаній, які часто здійснюють діяльність як на території України, так і за її межами, при зборі та використанні персональних даних необхідно враховувати всі відповідні законодавчі вимоги. Це є особливо актуальним для компаній, що обслуговують клієнтів або мають співробітників у різних країнах, адже вони повинні дотримуватися не лише українського законодавства про захист персональних даних, але й вимог, встановлених у юрисдикціях їхньої діяльності. Наприклад, якщо компанія веде діяльність в Україні та США, то зобов'язана враховувати положення Загального регламенту про захист даних (GDPR) при роботі з даними громадян Європейського Союзу, а також вимоги Каліфорнійського закону про захист прав споживачів (CCPA) при обробці даних резидентів Каліфорнії. Це вимагає від ІТ-компаній розробки комплексних політик та процедур, які відповідали б різноманітним регуляторним вимогам, а також забезпечення ефективного моніторингу та аудиту своїх процесів збору та обробки даних.

Законодавчі вимоги та положення щодо захисту персональних даних становлять основу для розробки політик конфіденційності в організаціях. Ці політики визначають правила та процедури збору, зберігання, обробки та передачі персональних даних, спрямовані на мінімізацію ризиків витоку інформації і відповідність законодавчим нормам.

Основні аспекти політик конфіденційності включають визначення цілей збору та використання персональних даних, а також типів зібраних даних для забезпечення розуміння та прозорості всіма сторонами. Політика описує методи та інструменти збору даних, такі як автоматизовані системи та опитувальні форми, здійснюючи збір відповідно до законодавчих норм. Ключовим аспектом є отримання згоди осіб на обробку їх персональних даних, що вимагає чіткої процедури її отримання, зберігання та можливості відкликання цієї згоди. Важливою складовою політики є заходи щодо збереження та захисту даних від несанкціонованого доступу або втрати, які можуть включати застосування шифрування, забезпечення надійності паролів та регулярне оновлення програмного забезпечення. Політика також має інформувати осіб про їх права,

включаючи право на доступ до власних персональних даних, виправлення, видалення, обмеження обробки та перенесення їх даних. У випадку передачі даних третім сторонам умови цієї передачі та заходи, які забезпечують конфіденційність цих даних повинні бути чітко описані [14]. Політики конфіденційності не тільки забезпечують захист персональних даних, але й сприяють підвищенню довіри співробітників та клієнтів до організації. Вони є ключовим інструментом управління ризиками та забезпечують дотримання вимог законодавства, що регулює обробку персональних даних.

Для успішної імплементації законодавчих норм щодо захисту персональних даних і забезпечення відповідності політик конфіденційності, організації використовують різноманітні технічні засоби та інструменти. Ці засоби допомагають гарантувати безпеку даних на всіх етапах їх життєвого циклу, від збору до знищення, з метою мінімізації ризиків і забезпечення конфіденційності інформації. Такі засоби включають в себе застосування шифрування для захисту даних під час їх передачі та зберігання. Також вони охоплюють впровадження систем аутентифікації, наприклад двофакторної, щоб забезпечити контроль доступу до інформації. Для захисту мережі використовуються брандмауери та віртуальні приватні мережі (VPN). Організації також розробляють системи для моніторингу безпекових інцидентів та проводять регулярні аудити для виявлення можливих загроз і вразливостей [32].

Забезпечення якості даних є невід'ємною частиною процесу застосування аналітики даних у сфері управління персоналом. Висока якість даних є критично важливою для достовірних аналітичних висновків та ефективного управління ресурсами підприємства. Основні аспекти забезпечення якості даних включають методи перевірки достовірності і цілісності даних, стратегії мінімізації помилок та уникнення неповноти і регулярний аудит та оновлення даних.

Перш за все, методи перевірки достовірності та цілісності даних є важливим етапом у процесі забезпечення їхньої якості. Ці методи включають перевірку на наявність аномалій та помилок, а також валідацію даних згідно з

встановленими стандартами та вимогами. Аномалії можуть включати неправильні значення, дублікати або несподівані варіації даних, які виходять за рамки очікуваного діапазону. Для виявлення таких аномалій застосовуються методи статистичного аналізу, порівняння з історичними даними або автоматизовані алгоритми, що виявляють відхилення. Перевірка на наявність помилок охоплює перевірку точності та достовірності даних, включно із правильністю форматування та відповідністю допустимим значенням параметрів [16]. Окрім цього, проведення валідації даних за встановленими стандартами і вимогами є необхідним для забезпечення відповідності специфічним правилам обробки та збереження даних, що визначаються для конкретного типу інформації.

Другим важливим аспектом є стратегії мінімізації помилок та уникнення неповноти даних, спрямовані на досягнення високої точності та повноти інформації. Наприклад, автоматизація процесів збору даних у форматі електронних анкет для реєстрації кадрових даних сприяє уникненню помилок, які можуть виникнути при ручному введенні інформації [31]. Крім того, використання інтегрованих систем управління персоналом дозволяє автоматизувати процеси збирання та зберігання даних про працівників, забезпечуючи їхню консистентність та актуальність без ризику втрати або фрагментарності інформації.

Крім того, систематичний аудит та періодичне оновлення даних є важливими складовими процесу забезпечення актуальності та відповідності поточним стандартам у сфері управління персоналом. Ці процедури сприяють збереженню високої якості даних на тривалий період, забезпечуючи їх готовність для ефективного використання у всіх аспектах управління персоналом. Наприклад, регулярний аудит персональних даних працівників дозволяє перевіряти їхню точність та актуальність, аби забезпечити, що інформація про кваліфікації, виконані проекти та навички завжди відповідає реальному стану справ [10]. Оновлення даних про навчання і розвиток

співробітників також є необхідним для підтримки їхнього професійного зростання та відповідності організаційним потребам.

Таким чином, забезпечення якості даних в контексті аналітики даних у сфері управління персоналом є важливою передумовою для успішного впровадження аналітичних інструментів та прийняття обґрунтованих управлінських рішень.

1.5. Формулювання завдання дослідження

Аналіз існуючих наукових робіт у сфері Data Science для HR та рекрутингу підтверджує, що впровадження аналітичних методів значно покращує ефективність процесів управління персоналом. Сучасні дослідження демонструють, що алгоритми машинного навчання активно використовуються для аналізу резюме, оцінки компетенцій кандидатів, автоматизації відбору персоналу та прогнозування їхньої майбутньої ефективності. Використання таких підходів дозволяє компаніям ухвалювати обґрунтовані кадрові рішення, скорочуючи час найму та зменшуючи витрати на рекрутинг.

Проте, незважаючи на значний прогрес у застосуванні предиктивної аналітики для оцінки кандидатів, значно менше уваги приділено питанням прогнозування термінів закриття вакансій. Існуючі дослідження зосереджені переважно на визначенні відповідності кандидатів вакансії, оцінці їхніх компетенцій та потенціалу, але при цьому залишають поза увагою питання ефективного планування самого процесу найму.

Зокрема, моделі на кшталт Random Forest та Gradient Boosting вже показали свою ефективність у відборі кандидатів [2], а використання графових нейронних мереж (GNNs) дозволило покращити аналіз взаємозв’язків між кандидатами [9]. Аналіз ключових факторів, що впливають на успішність кандидатів, здійснювався через регресійні моделі та алгоритми класифікації [22], що підтвердило вплив технічних і поведінкових навичок на конкурентоспроможність працівників. Проте навіть у масштабному дослідженні

Пессаха, яке використовує величезний набір історичних рекрутингових даних, фокус робиться на оптимізації процесу найму [28], але не на передбаченні часових рамок його реалізації.

Оскільки ефективне управління людськими ресурсами потребує не лише якісного відбору кандидатів, а й точного прогнозування термінів закриття вакансій, ця проблема стає критичною для організацій. Затримки в наймі можуть призводити до втрати продуктивності, додаткових фінансових витрат та порушення бізнес-процесів. Успішне прогнозування термінів закриття вакансій дозволило б компаніям ефективніше розподіляти ресурси, визначати оптимальні канали залучення кандидатів та коригувати HR-стратегії.

Застосування методів Data Science для розв'язання цієї проблеми є перспективним напрямом досліджень. Використання предиктивної аналітики на основі історичних HR-даних може допомогти створити модель прогнозування термінів закриття вакансій, яка враховуватиме:

- Тип вакансії (посада, рівень складності, вимоги до кандидата)
- Джерело пошуку кандидатів (внутрішній найм, рекрутингові платформи, рекомендації тощо)
- Попередній досвід компанії у заповненні аналогічних позицій
- Конкурентне середовище (ситуація на ринку праці, рівень зарплат)
- Загальну статистику рекрутингових процесів компанії

Таким чином, головний виклик цього дослідження – **розробка моделі**, яка дозволить HR-відділам **прогнозувати часові рамки закриття вакансій** з високою точністю, **використовуючи методи машинного навчання**. Це не тільки підвищить ефективність рекрутингу, а й забезпечить оптимізацію кадрових стратегій у довготривалій перспективі.

Отже, у рамках даного дослідження пропонується розробка аналітичної моделі прогнозування термінів закриття вакансій, що дозволить компаніям:

- Оптимізувати процеси рекрутингу, скорочуючи невизначеність у наймі.

- Підвищити ефективність використання ресурсів, прогнозуючи затрати часу та необхідність залучення додаткових HR-інструментів.
- Підготувати HR-аналітику до гнучкого реагування на зміни ринку праці, що сприятиме стратегічному плануванню найму в організаціях.

Відповідно до сформульованої мети, у дослідженні поставлено такі завдання:

- Проаналізувати теоретичні та практичні аспекти використання Data Science в управлінні персоналом.
- Визначити та обрати моделі і методи Data Science для вирішення визначених проблем.
- Реалізувати обрані моделі та методи на мові програмування R.
- Провести аналіз отриманих результатів та оцінити ефективність використаних моделей у HR-процесах.

Висновки

Цей розділ підтверджує зростаючу значущість Data Science як інструменту стратегічного управління у сфері HR. Теоретичний огляд засвідчив, що наука про дані на сучасному етапі є ключовим елементом цифрової трансформації організацій. Завдяки інтеграції статистичних методів, машинного навчання та інструментів візуалізації, Data Science дозволяє управлінню персоналом перейти від інтуїтивного управління до моделі, орієнтованої на обґрунтовані даними рішення.

Огляд наукових публікацій доводить, що найбільш активно Data Science використовується для аналізу ефективності рекрутингових кампаній, прогнозування відтоку персоналу, а також оцінки відповідності кандидатів на етапі підбору. Водночас встановлено, що тематика прогнозування строків закриття вакансій, попри свою практичну важливість, залишається недостатньо дослідженою. Це створює обґрунтовану нішу для наукового пошуку.

Додатково в розділі проаналізовано типові джерела HR-даних, що становлять основу для побудови моделей, а також розглянуто підходи до забезпечення їх якості та дотримання вимог конфіденційності. Зокрема, окреслено ключові юридичні засади (в тому числі відповідність GDPR), технічні та організаційні механізми, необхідні для захисту персональної інформації співробітників. Увага до цих аспектів є передумовою як достовірності аналітичних висновків, так і довіри до HR-аналітики з боку працівників та стейкхолдерів.

Теоретичний і літературний аналіз, проведений у межах розділу, дозволяє зробити висновок про обґрунтованість та актуальність поставлених у вступі завдань дослідження. Виявлена у наукових джерелах фрагментарність у використанні методів Data Science у сфері HR підкріплює необхідність системного теоретичного й прикладного вивчення цієї проблематики. Недостатня увага до прогнозування термінів закриття вакансій свідчить про потребу у застосуванні відповідних аналітичних підходів та моделей машинного навчання. Аналіз джерел HR-даних і вимог до їхньої якості створює основу для подальшої реалізації моделі засобами програмного забезпечення, а також обґрунтовує важливість ретельної оцінки її ефективності. Таким чином, результати першого розділу формують логічне підґрунтя для реалізації всіх ключових завдань дослідження, пов’язаних з вибором методології, побудовою прогнозної моделі та перевіркою її прикладної цінності у сфері управління персоналом.

РОЗДІЛ 2
ФОРМАЛІЗАЦІЯ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ ТЕРМІНІВ ЗАКРИТТЯ ВАКАНСІЙ ІЗ ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ DATA SCIENCE

2.1. Визначення даних для проведення наукового дослідження

Ефективне прогнозування часу закриття вакансій є важливим завданням для HR-аналітики, оскільки дозволяє компаніям оптимізувати процес рекрутингу, розподіл ресурсів та стратегічне планування персоналу. Одним із ключових факторів для побудови точної та надійної моделі прогнозування є використання якісних вихідних даних. У даному дослідженні для аналізу використано синтетично згенеровані дані, побудовані на основі реальних даних бази практики, отриманих із системи управління людськими ресурсами (HRIS) BambooHR. Синтетичні дані максимально наближені до реальних за структурою та основними характеристиками, що дозволяє забезпечити об'єктивність дослідження та відповідність умовам реального рекрутингового процесу.

BambooHR є автоматизованою платформою для управління персоналом, що використовується для збору, збереження та аналізу інформації про співробітників та рекрутингові процеси. Використання такої системи дозволяє отримати структуровані та стандартизовані HR-дані, що є необхідним для проведення якісного аналізу. На відміну від ручного введення даних або суб’єктивних оцінок рекрутерів, HRIS забезпечує об’єктивність і точність інформації, що підвищує достовірність результатів дослідження.

Для побудови моделі прогнозування термінів закриття вакансій було сформовано синтетичний набір даних, який відтворює реальні рекрутингові дані компанії за останні три роки, починаючи з кінця 2021 року, коли у компанії було впроваджено систему HRIS BambooHR. Вибір трирічного періоду обумовлений моментом запровадження єдиної системи збору HR-даних, що забезпечує їхню структурованість і порівнянність у межах аналізованого періоду. Така тривалість

спостереження також дозволяє охопити зміну рекрутингових процесів, сезонні коливання та загальні тенденції ринку праці.

Вибір метрик для моделювання базується на аналізі інформації, отриманої від HR-відділу компанії, щодо основних факторів, які, на їхню думку, мають найбільший вплив на швидкість закриття вакансій. У процесі збору даних було визначено, що ключовими змінними є характеристики самої вакансії, зокрема її категорія, рівень досвіду та заробітна плата, оскільки вони визначають складність пошуку відповідних кандидатів. Також було враховано параметри, що описують процес найму, зокрема кількість кандидатів і кількість етапів відбору, адже ці показники безпосередньо впливають на тривалість рекрутингового циклу. Окрему увагу приділено джерелам пошуку кандидатів, оскільки ефективність різних каналів найму може суттєво відрізнятися. Крім того, було визначено, що сезонність також відіграє роль у динаміці рекрутингових процесів, адже в певні періоди року активність пошукачів змінюється, що впливає на загальну тривалість найму. Усі ці метрики були відібрані з урахуванням їхньої доступності в HRIS (BambooHR) та можливості подальшого використання у прогнозних моделях без необхідності залучення суб'єктивних оцінок. Таким чином, сформований набір змінних дозволяє всебічно охопити основні аспекти процесу закриття вакансій та забезпечити достовірність отриманих прогнозів.

Дані, представлені в датасеті, можна розподілити на кілька основних категорій:

1. Характеристики вакансії
   - *Назва вакансії* – відображає професійну спрямованість посади, що дозволяє класифікувати її за напрямом діяльності та виявляти потенційні відмінності у швидкості закриття різних типів вакансій.
   - *Категорія вакансії* – узагальнена класифікація, що відображає належність вакансії до певної сфери діяльності компанії (наприклад, розробка програмного забезпечення, аналітика, маркетинг, HR).

- *Рівень досвіду* – позначає кваліфікаційні вимоги до кандидата (Junior, Middle, Senior), що є критичним фактором, оскільки рівень складності позиції впливає на швидкість пошуку відповідного спеціаліста.
- *Заробітна плата* – відображає рівень матеріальної компенсації, що є одним із ключових факторів привабливості вакансії на ринку праці.

2. Процес закриття вакансії

- *Дата відкриття та дата закриття вакансії* – використовуються для обчислення цільової змінної дослідження – кількості днів, необхідних для заповнення вакансії.
- *Кількість кандидатів* – показник загальної кількості осіб, які подали заявку або були розглянуті на відповідну вакансію, що дає змогу оцінити рівень попиту на дану посаду.
- *Кількість етапів відбору* – параметр, що характеризує тривалість процесу найму, оскільки вакансії з багаторівневими етапами оцінювання (наприклад, тестові завдання, технічні співбесіди, фінальні інтерв'ю) можуть вимагати додаткового часу для прийняття рішення.

3. Канали пошуку кандидатів

- *Джерело кандидатів* – вказує, через який канал було знайдено претендента (наприклад, внутрішній найм, рекомендації, рекрутингові агентства, онлайн-платформи, зокрема LinkedIn), що дозволяє оцінити ефективність кожного джерела у контексті швидкості закриття вакансій.

4. Зовнішні фактори

- *Сезонність* – характеристика, що позначає період року, у який була закрита вакансія (зима, весна, літо, осінь). Врахування сезонних змін дає можливість визначити потенційні коливання у швидкості закриття вакансій, що може бути пов'язано з поведінковими патернами кандидатів та загальною динамікою ринку праці.

Сформований набір даних містить об'єктивні показники, які безпосередньо впливають на тривалість процесу закриття вакансій, що забезпечує надійність і практичну значущість отриманих прогнозів. Обрані

змінні охоплюють як внутрішні параметри вакансій та процесу рекрутингу, так і зовнішні фактори, що дає змогу отримати комплексний аналіз залежностей між різними характеристиками найму. Така структура даних сприяє побудові високоточних моделей прогнозування, здатних враховувати різноманітні аспекти, що впливають на швидкість закриття вакансій.

2.2 Вибір методології для вирішення задачі прогнозування термінів закриття вакансій

Основним завданням даного дослідження є побудова прогнозної моделі для оцінки термінів закриття вакансій на основі рекрутингових даних. Реалізація такого прогнозування дозволяє підвищити ефективність процесу найму шляхом покращення планування кадрових ресурсів та оптимізації роботи рекрутингових команд.

З огляду на характер прогнозованої змінної – кількість днів до закриття вакансії – завдання моделювання належить до завдань регресії в машинному навчанні. Вхідні дані для аналізу включають як числові (наприклад, заробітна плата, кількість кандидатів), так і категоріальні ознаки (наприклад, категорія вакансії, джерело кандидатів), що вимагає від обраної моделі здатності ефективно працювати з різнотипними змінними. Крім того, у даних можуть бути присутні складні нелінійні взаємозв’язки між ознаками, що потребує використання методів, здатних їх враховувати.

У зв’язку з цим, для вирішення поставленого завдання необхідно обрати таку модель машинного навчання, яка забезпечує високу точність прогнозування, стійкість до перенавчання, ефективну обробку як числових, так і категоріальних ознак, а також дозволяє інтерпретувати важливість факторів, що впливають на результати прогнозу. З огляду на специфіку завдання прогнозування термінів закриття вакансій, доцільним є розгляд сучасних методів машинного навчання, що широко застосовуються для аналізу числових показників. До таких методів належать лінійна регресія, дерева рішень,

ансамблеві алгоритми (зокрема Random Forest та Gradient Boosting), метод опорних векторів для регресії та нейронні мережі.

Лінійна регресія є одним із базових методів моделювання залежностей між числовими змінними у задачах прогнозування. Вона передбачає побудову найкращої лінійної апроксимації між незалежними змінними та цільовою змінною шляхом мінімізації суми квадратів помилок [18].

Формула лінійної регресії у математичному записі:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon, \text{ де} \qquad (1)$$

$\hat{y}$ – прогнозоване значення цільової змінної;

$\beta_0, \beta_1, \ldots \beta_p$ – коефіцієнти моделі;

ε – випадкова похибка.

Основними перевагами лінійної регресії є простота реалізації, швидкість навчання та легкість інтерпретації результатів. Водночас метод має обмеження у випадках складних або нелінійних залежностей між змінними, що може впливати на точність прогнозування.

Дерева рішень є одним із методів прогнозування числових змінних, що дозволяють моделювати залежність між вхідними ознаками та цільовою змінною шляхом рекурсивного розбиття простору ознак на підмножини [18]. У задачах регресії дерево рішень будується таким чином, щоб у кожному вузлі мінімізувати середню квадратичну помилку прогнозу. Формально цей критерій має вигляд:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \text{ де} \qquad (2)$$

y_{i-} фактичне значення цільової змінної;

$\hat{y}_i -$ прогнозоване значення;

$n -$ кількість спостережень у вузлі.

Основними перевагами дерев рішень є їхня здатність враховувати складні нелінійні взаємозв’язки між ознаками, працювати з числовими та категоріальними даними без необхідності попередньої обробки, а також зрозуміла інтерпретація результатів. Попри таку здатність, цей метод має суттєве

обмеження – високу схильність до перенавчання, особливо при побудові глибоких дерев. Через це окремі дерева рішень часто демонструють хорошу якість на навчальній вибірці, але погіршують точність прогнозів на нових, невідомих даних. Для подолання цього недоліку в задачах прогнозування все ширше застосовуються ансамблеві методи, які комбінують результати декількох моделей задля підвищення стабільності та точності прогнозів.

Серед найбільш ефективних ансамблевих підходів варто виокремити алгоритм *Random Forest*, який базується на побудові великої кількості дерев рішень, кожне з яких навчається на випадковій підмножині спостережень і ознак, що дозволяє суттєво знизити варіацію окремих моделей і підвищити стабільність прогнозів. Однією з ключових переваг Random Forest є його стійкість до перенавчання, оскільки усереднення результатів багатьох слабких моделей забезпечує краще узагальнення на нових даних [4]. Крім того, Random Forest здатний працювати як із числовими, так і з категоріальними змінними без необхідності складної попередньої обробки даних. Завдяки цим властивостям Random Forest може бути ефективним інструментом для вирішення задач прогнозування у сфері HR-аналітики, зокрема для моделювання термінів закриття вакансій. Формально принцип усереднення прогнозів у Random Forest можна подати у вигляді такої формули:

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x), \text{ де} \qquad (3)$$

$\hat{f}(x) -$ підсумковий прогноз ансамблю для спостереження x;

$B -$ кількість дерев у моделі;

$T_b(x) -$ прогноз окремого дерева.

Ансамблеві методи, зокрема Random Forest і Gradient Boosting, поєднують результати багатьох базових моделей, що дозволяє підвищити точність прогнозування та зменшити ризик перенавчання. Random Forest характеризується стійкістю до шумів і високою стабільністю результатів, тоді як *Gradient Boosting* дозволяє досягати високої точності за рахунок послідовного вдосконалення помилок попередніх моделей. Механізм оновлення моделі на кожній ітерації описується наступною формулою:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \text{ де} \quad (4)$$

$F_m(x)$ – поточне передбачення ансамблю;

$h_m(x)$ – нова базова модель, що коригує помилки;

γ_m – коефіцієнт навчання.

На кожній ітерації нова модель фокусується на тих об'єктах, які були погано передбачені попередньою моделлю, що дозволяє поступово зменшувати загальну похибку системи [25]. Такий підхід забезпечує високу гнучкість моделі та дозволяє ефективно моделювати складні, нелінійні залежності у даних. Однак Gradient Boosting є більш чутливим до налаштування гіперпараметрів та може бути схильним до перенавчання за відсутності відповідної регуляризації.

Метод опорних векторів для регресії (SVR) є ефективним підходом для прогнозування числових змінних, який базується на побудові функції наближення, що мінімізує відхилення прогнозованих значень від фактичних у межах заданого порогу чутливості. Основною перевагою SVR є здатність моделювати складні залежності навіть при обмеженій кількості навчальних даних за рахунок спеціального математичного перетворення ознак, яке дозволяє відобразити дані у такій формі, де їх взаємозв'язки стають простішими для прогнозування [18]. Прогнозування у SVR здійснюється за допомогою ядрової функції, що оцінює схожість між новим спостереженням та опорними векторами:

$$\hat{y}(x) = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i, x) + b, \text{ де} \quad (5)$$

$(a_i - a_i^*)$ – параметри, визначені під час навчання;

$K(x_i, x)$ – ядрова функція, яка перетворює простір ознак;

x_i – опорні вектори;

b – зсув.

Модель демонструє високу стійкість до перенавчання та здатність зберігати високу точність прогнозування, що робить її придатною для задач HR-аналітики, пов’язаних із прогнозуванням термінів закриття вакансій в умовах обмеженого обсягу даних, однак модель потребує ретельного налаштування гіперпараметрів.

Нейронні мережі демонструють високу ефективність у задачах з великими обсягами даних та складною структурою залежностей. Обчислення вихідного сигналу в одному шарі мережі зазвичай ґрунтується на ваговому перетворенні вхідних ознак із подальшим застосуванням активаційної функції:

$$\hat{y} = \sigma(\sum_{i=1}^{n} w_i x_i + b, \text{ де} \quad (6)$$

x_i – значення вхідних ознак;

w_i – вагові коефіцієнти;

b – зсув;

σ – функція активації (наприклад, ReLU або сигмоїда).

Завдяки багатошаровій архітектурі вони здатні автоматично виявляти приховані закономірності у великих обсягах даних, що забезпечує високу гнучкість моделювання [11]. Однак ефективне навчання нейронних мереж потребує значної кількості даних і ретельного налаштування гіперпараметрів. За умови обмежених обсягів даних, що часто спостерігається у практичних HR-задачах, використання нейронних мереж може призводити до перенавчання та нестабільності результатів. Ці фактори обмежують доцільність застосування нейронних мереж для прогнозування термінів закриття вакансій у випадках невеликих вибірок.

Завдання прогнозування термінів закриття вакансій передбачає вибір моделей, здатних обробляти різнотипні дані та враховувати складні взаємозв’язки між ознаками. Лінійна регресія забезпечує простоту й інтерпретованість, проте обмежена при моделюванні нелінійних залежностей. Дерева рішень і ансамблеві методи, зокрема Random Forest і Gradient Boosting, дозволяють ефективно працювати з гетерогенними даними і складними структурами, хоча Gradient Boosting вимагає ретельного налаштування для уникнення перенавчання. Метод опорних векторів для регресії демонструє хорошу точність навіть на малих вибірках, але потребує оптимізації параметрів. Нейронні мережі відзначаються високою гнучкістю при великих обсягах даних, однак у випадку обмежених вибірок їх застосування пов'язане з ризиком перенавчання та нестабільністю прогнозів.

2.3 Застосування методу Random Forest для прогнозування термінів закриття вакансій: переваги та обмеження

Для вирішення завдання прогнозування термінів закриття вакансій було обрано алгоритм Random Forest, який є одним із найефективніших методів машинного навчання для роботи з гетерогенними даними. Його застосування дозволяє отримати точні прогнози, враховуючи вплив різних факторів, таких як рівень досвіду кандидата, джерело пошуку, кількість етапів відбору, розмір заробітної плати тощо.

Основні причини використання Random Forest у цьому дослідженні:

1. Гарно працює з числовими та категоріальними змінними
   - Набір даних включає числові (заробітна плата, кількість кандидатів, кількість етапів відбору) та категоріальні змінні (рівень досвіду, джерело кандидата, сезон закриття вакансії). Random Forest здатний ефективно обробляти такі ознаки без складних процедур попередньої трансформації.
2. Стійкість до викидів та нерівномірності розподілу
   - У HR-аналітиці нерідко спостерігаються суттєві відмінності між вакансіями залежно від рівня посади чи специфіки вимог. Random Forest є менш чутливим до аномальних спостережень у даних порівняно з іншими моделями, такими як лінійна регресія.
3. Автоматичне визначення важливості змінних
   - Метод дозволяє визначити, які фактори найбільше впливають на швидкість закриття вакансій, що є корисним для прийняття управлінських рішень у HR-аналітиці.
4. Висока точність прогнозів
   - Алгоритм використовує комбінацію великої кількості дерев рішень, що значно покращує стійкість до перенавчання та збільшує точність порівняно з одиночними моделями.

5. Не потребує нормалізації та масштабування даних

- Оскільки метод працює з бінарними розгалуженнями, він не вимагає перетворення змінних у єдиний масштаб, що спрощує підготовку даних.

Потенційні обмеження Random Forest:

Хоча метод має багато переваг, слід враховувати такі аспекти:

- Відсутність простої інтерпретованої формули: у випадку, коли потрібно отримати чітке рівняння для розрахунку термінів закриття вакансії, простіші моделі (наприклад, лінійна регресія) можуть бути більш підходящими.
- Високе споживання ресурсів при масштабуванні: хоча в даному випадку набір даних є відносно невеликим, на великих обсягах Random Forest може потребувати більше часу для навчання.

Враховуючи виявлені переваги та обмеження, застосування методу Random Forest дозволяє побудувати надійну прогнозну модель, адаптовану до специфіки рекрутингових даних, що створює підґрунтя для подальшого етапу практичної реалізації дослідження.

Враховуючи обґрунтованість вибору алгоритму Random Forest для вирішення поставленого завдання, наступним кроком є визначення критеріїв оцінки його ефективності. У задачах регресійного прогнозування, зокрема таких, що стосуються оцінки тривалості закриття вакансій, доцільно використовувати метрики, які вимірюють точність передбачених значень у безпосередніх одиницях цільової змінної. Найпоширенішими серед них є середня абсолютна похибка (MAE) та корінь з середньої квадратичної похибки (RMSE).

Метрика **MAE (Mean Absolute Error)** обчислює середнє абсолютне відхилення між прогнозованими та фактичними значеннями, забезпечуючи інтуїтивну інтерпретацію у тих самих одиницях, що й цільова змінна (у цьому випадку – дні). Формула має вигляд:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \text{ де} \qquad (7)$$

y_i – фактичне значення;

$\hat{y}_i$ – прогнозоване значення;

n – кількість спостережень.

Метрика **RMSE (Root Mean Square Error)**, натомість, акцентує на великих відхиленнях, підносячи похибки до квадрату перед усередненням:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (8)$$

Застосування обох метрик дозволяє оцінити не лише середній рівень похибки моделі, а й її чутливість до екстремальних відхилень. Вибір цих показників є обґрунтованим у контексті HR-аналітики, оскільки дає змогу точно інтерпретувати помилки прогнозування термінів закриття вакансій та оцінити практичну корисність моделі.

2.4 Мова програмування, основні бібліотеки та інструменти

У процесі розробки моделі прогнозування термінів закриття вакансій використовувалася мова програмування R, яка є одним із провідних інструментів для аналізу даних і побудови моделей машинного навчання. Вибір R був зумовлений її широкими можливостями для статистичного моделювання, розвиненою екосистемою пакетів для роботи з табличними даними та гнучкістю у візуалізації результатів.

Основні етапи обробки даних і побудови моделі здійснювалися із використанням таких бібліотек:

- readxl — для імпорту даних із формату .xlsx у середовище R;
- dplyr — для обробки і трансформації даних, включаючи групування, агрегацію та фільтрацію спостережень;
- ggplot2 — для побудови графічної візуалізації результатів дослідження, зокрема аналізу залежностей між змінними;
- randomForest — для побудови прогнозної моделі методом ансамблю дерев рішень;

- caret — для розбиття даних на навчальну і тестову вибірки, а також для оптимізації параметрів моделі за допомогою крос-валідації;
- tidyr — для структуризації даних перед моделюванням.

Для обробки даних та побудови моделі застосовувався комплексний підхід, що включав попередню очистку датасету, перетворення категоріальних змінних у відповідний формат (фактори), видалення нерелевантних змінних, а також розробку стратегії стратифікованого розбиття вибірки для уникнення дисбалансу класів.

Завдяки використанню зазначених інструментів було забезпечено ефективне виконання всіх етапів дослідження: від аналізу та візуалізації даних до розробки, налаштування та оцінки прогнозної моделі.

Висновки

У другому розділі було здійснено комплексну формалізацію методології побудови прогнозної моделі для оцінки термінів закриття вакансій на основі технологій Data Science. Проведено обґрунтування вибору джерел даних, методів моделювання, алгоритмів машинного навчання та інструментів програмної реалізації. Узагальнення результатів дозволяє зробити низку важливих висновків, що становлять методологічне підґрунтя для подальшої практичної реалізації.

Перш за все, визначено доцільність використання синтетичних даних, згенерованих на основі реальних записів із HRIS-системи BambooHR. Такий підхід забезпечив збереження структурної відповідності реальному бізнес-контексту при дотриманні вимог конфіденційності. Структура сформованого датасету охоплює ключові аспекти рекрутингового процесу: характеристики вакансії, параметри відбору, джерела кандидатів і зовнішні сезонні чинники. Це дозволяє забезпечити комплексний аналіз і високий рівень прогностичної точності.

У межах методологічного аналізу порівняно низку моделей машинного навчання, що можуть бути застосовані для вирішення задачі регресії. Було з’ясовано, що метод Random Forest є оптимальним вибором для поставленої задачі з огляду на його здатність ефективно обробляти як числові, так і категоріальні змінні, стійкість до перенавчання, а також наявність механізмів оцінки важливості предикторів. Попри деякі обмеження, зокрема щодо складності інтерпретації та ресурсоємності, Random Forest забезпечує високу стабільність результатів, що критично важливо у сфері HR-аналітики.

Також визначено набір програмних засобів, що забезпечують реалізацію усіх етапів дослідження: від імпорту та попередньої обробки даних до побудови й оцінки моделі. Обґрунтовано використання мови програмування R та спеціалізованих бібліотек, таких як randomForest, caret, ggplot2, що дозволяють не лише реалізувати алгоритмічну частину моделі, але й забезпечити її візуалізацію та інтерпретацію.

Таким чином, викладена в розділі методологія дозволяє забезпечити належний рівень обґрунтованості прогнозування тривалості закриття вакансій, а також створює основу для подальшого впровадження моделі в прикладне середовище HR-аналітики.

РОЗДІЛ 3
РОЗРОБКА МЕТОДУ DATA SCIENCE ДЛЯ ПРОГНОЗУВАННЯ ТЕРМІНІВ ЗАКРИТТЯ ВАКАНСІЙ

3.1 Розробка технології рішення науково-прикладної задачі

Після проведення аналізу предметної області, формулювання дослідницького завдання та обґрунтування вибору методів машинного навчання, наступним етапом дослідження є розробка прикладного рішення на основі технологій Data Science. Основною метою цього етапу є побудова, налаштування та оцінка прогнозної моделі, здатної передбачити терміни закриття вакансій на основі доступних рекрутингових даних. Така модель може стати ефективним інструментом для HR-аналітики, що дозволяє приймати обґрунтовані рішення щодо оптимізації процесу найму персоналу.

Процес побудови прогнозної моделі у сфері HR-аналітики вимагає чіткої структуризації етапів, які охоплюють підготовку даних, первинний аналіз, виявлення закономірностей, а також технічне забезпечення реалізації моделі. Метою цього етапу є перетворення вихідного масиву рекрутингових даних у формалізований набір ознак, що можуть бути використані для побудови регресійної моделі, здатної прогнозувати терміни закриття вакансій.

Розробка технології рішення задачі прогнозування термінів закриття вакансій розпочинається з імпорту даних та їх попереднього аналізу. На цьому етапі здійснюється завантаження синтетично згенерованого датасету, створеного на основі типових рекрутингових характеристик, що спостерігалися в реальному середовищі. Дані зберігаються у форматі Excel, що дозволяє зручно імпортувати їх до середовища R за допомогою пакету readxl. Метою є перевірка структури даних, оцінка наявності пропущених значень, виявлення типів змінних, а також первинна підготовка інформації до подальшого аналітичного опрацювання.

Після імпорту даних здійснимо попередню структурну діагностику таблиці спостережень. Для цього використаємо функції summary() та str(), які

дозволяють сформувати загальне уявлення про типи змінних, розподіл значень, а також виявити потенційні пропущені або аномальні значення. Такий аналіз є необхідною передумовою для прийняття подальших рішень щодо очищення та трансформації даних.

```
> str(data)
tibble [91 × 11] (S3: tbl_df/tbl/data.frame)
 $ job_title       : chr [1:91] "Business Analyst" "QA Engineer" "Software Engineer"
"Technical Support Enginneer" ...
 $ category        : chr [1:91] "Analytics" "Development" "Development" "Tech Support"
...
 $ experience_level: chr [1:91] "Middle" "Senior" "Junior" "Senior" ...
 $ salary          : num [1:91] 2100 3900 1000 2700 1600 7000 2300 2850 4200 1600 ...
 $ open_date       : POSIXct[1:91], format: "2024-12-25" "2024-12-01" ...
 $ close_date      : POSIXct[1:91], format: "2025-01-31" "2025-01-05" ...
 $ days_to_close   : num [1:91] 37 35 21 44 41 62 24 30 49 20 ...
 $ num_candidates  : num [1:91] 42 54 89 37 43 25 19 36 28 15 ...
 $ hiring_steps    : num [1:91] 3 4 4 3 3 4 4 3 4 3 ...
 $ source          : chr [1:91] "LinkedIn" "Recommendation" "Dou" "LinkedIn" ...
 $ season          : chr [1:91] "winter" "winter" "winter" "winter" ...
```

Рисунок 1 – Відображення функції str.

Як видно з результату застосування функції str(), об’єктом є таблиця типу tibble з 91 спостереженням та 11 змінними. Ключові змінні salary, days_to_close, num_candidates та hiring_steps мають числовий тип, що є придатним для регресійного аналізу. Змінні job_title, category, experience_level, source та season зчитані як character і потребують перетворення у фактори для коректної обробки у моделі. Змінні open_date та close_date представлені у форматі POSIXct, що дозволяє здійснювати хронологічні обчислення. Таким чином, попередня діагностика типів підтверджує відповідність структури даних вимогам до побудови прогнозної моделі та визначає необхідні кроки для подальшої трансформації змінних.

```
> summary(data)
  job_title           category         experience_level       salary
 Length:91          Length:91          Length:91          Min.   : 600
 Class :character   Class :character   Class :character   1st Qu.:1750
 Mode  :character   Mode  :character   Mode  :character   Median :2300
                                                          Mean   :2530
                                                          3rd Qu.:3000
                                                          Max.   :7000

   open_date                       close_date
 Min.   :2021-12-12 00:00:00.00   Min.   :2022-01-10 00:00:00.00
 1st Qu.:2022-10-12 12:00:00.00   1st Qu.:2022-10-26 12:00:00.00
 Median :2023-08-29 00:00:00.00   Median :2023-09-11 00:00:00.00
 Mean   :2023-07-26 09:45:29.67   Mean   :2023-08-17 20:50:06.58
 3rd Qu.:2024-07-14 12:00:00.00   3rd Qu.:2024-08-09 00:00:00.00
 Max.   :2025-01-04 00:00:00.00   Max.   :2025-01-31 00:00:00.00
 days_to_close   num_candidates    hiring_steps      source
 Min.   : 9.00   Min.   :  9.00   Min.   :2.000   Length:91
 1st Qu.:16.00   1st Qu.: 19.00   1st Qu.:2.000   Class :character
 Median :19.00   Median : 29.00   Median :3.000   Mode  :character
 Mean   :22.46   Mean   : 36.36   Mean   :2.879
 3rd Qu.:25.00   3rd Qu.: 48.50   3rd Qu.:3.000
 Max.   :64.00   Max.   :112.00   Max.   :4.000
    season
 Length:91
 Class :character
 Mode  :character
```

Рисунок 2 – Відображення функції summary.

Функція summary() надала базові описові характеристики змінних. Встановлено, що змінна salary має діапазон значень від 600 до 7000 одиниць, а медіанне значення становить 2300, що свідчить про помірно зсунутий розподіл. Ключова цільова змінна days_to_close має медіану 19 днів, із розкидом значень від 9 до 64. Також варто відзначити, що num_candidates варіюється від 9 до 112, а кількість етапів відбору не перевищує чотирьох. Отримані статистики підтверджують наявність достатньої варіативності ознак для побудови прогнозної моделі.

Наступним кроком є перевірка датасету на наявність пропущених значень, оскільки вони можуть суттєво впливати на якість моделювання та інтерпретацію результатів. Для цього використаємо функцію colSums(is.na(...)), яка дозволяє визначити кількість відсутніх спостережень у кожному стовпці. У разі виявлення незначної кількості пропусків (до 5% від загального обсягу даних), доцільно

вилучити відповідні записи, щоб уникнути викривлення структури змінних або введення штучно згенерованих значень.

```
> colSums(is.na(data))
      job_title        category experience_level          salary
              0               0               0               0
      open_date      close_date    days_to_close   num_candidates
              0               0               0               0
   hiring_steps          source          season
              0               0               0
```

Рисунок 3 – Відображення функції colSums(is.na(...)).

Бачимо, що результати перевірки засвідчили повну відсутність пропущених значень у всіх змінних. Це говорить про належну якість підготовки даних і дозволяє перейти до наступного етапу обробки без необхідності додаткових заходів щодо очищення.

Далі необхідно здійснити підготовку категоріальних змінних до аналізу. Зокрема, змінні experience_level, source та season, що представляють рівень досвіду кандидата, джерело пошуку та сезон найму відповідно, мають бути приведені до факторного типу. Це перетворення є критично важливим для забезпечення їх правильної інтерпретації алгоритмами машинного навчання, які розрізняють категоріальні та числові ознаки. Оскільки ці змінні мають дискретну природу та обмежений набір унікальних значень, їх кодування як факторів дозволить уникнути помилкових припущень про існування кількісної шкали та забезпечить коректне врахування їх впливу під час побудови прогнозної моделі.

```
> data$experience_level <- as.factor(data$experience_level)
> data$source <- as.factor(data$source)
> data$season <- as.factor(data$season)
```

Рисунок 4 – Перекодування змінних у фактори.

Застосуємо повторно функцію str() для перевірки успішності перекодування змінних у фактори:

```
> str(data)
tibble [91 × 11] (S3: tbl_df/tbl/data.frame)
 $ job_title       : chr [1:91] "Business Analyst" "QA Engineer" "Software Engineer"
"Technical Support Enginneer" ...
 $ category        : chr [1:91] "Analytics" "Development" "Development" "Tech Support"
...
 $ experience_level: Factor w/ 3 levels "Junior","Middle",..: 2 3 1 3 2 3 2 3 3 2 ...
 $ salary          : num [1:91] 2100 3900 1000 2700 1600 7000 2300 2850 4200 1600 ...
 $ open_date       : POSIXct[1:91], format: "2024-12-25" "2024-12-01" ...
 $ close_date      : POSIXct[1:91], format: "2025-01-31" "2025-01-05" ...
 $ days_to_close   : num [1:91] 37 35 21 44 41 62 24 30 49 20 ...
 $ num_candidates  : num [1:91] 42 54 89 37 43 25 19 36 28 15 ...
 $ hiring_steps    : num [1:91] 3 4 4 3 3 4 4 3 4 3 ...
 $ source          : Factor w/ 5 levels "Djinni","Dou",..: 3 5 2 3 3 3 3 3 3 2 ...
 $ season          : Factor w/ 4 levels "autumn","spring",..: 4 4 4 4 4 4 4 4 4 4 ...
```

Рисунок 5 – Відображення функції str.

Результати повторного виклику функції str() після трансформації змінних свідчать про успішне приведення трьох категоріальних змінних (experience_level, source, season) до факторного типу. Така трансформація є необхідною процедурою в рамках підготовки даних до моделювання із застосуванням алгоритму Random Forest, оскільки цей метод здатен коректно обробляти факторні змінні без потреби у їх попередньому перетворенні в числові або бінарні формати. Отримані рівні факторів адекватно відображають логічну структуру кожної ознаки: змінна experience_level має три рівні, що відповідають категоріям кваліфікації кандидатів, source класифікує джерела надходження кандидатів, а season – сезонні характеристики рекрутингових періодів.

Наступним кроком є оцінювання розподілу ключових змінних, зокрема цільової змінної days_to_close, що дозволяє виявити потенційні аномалії, асиметрію та варіативність у даних, а також сформувати попередні аналітичні припущення щодо структури рекрутингових процесів.

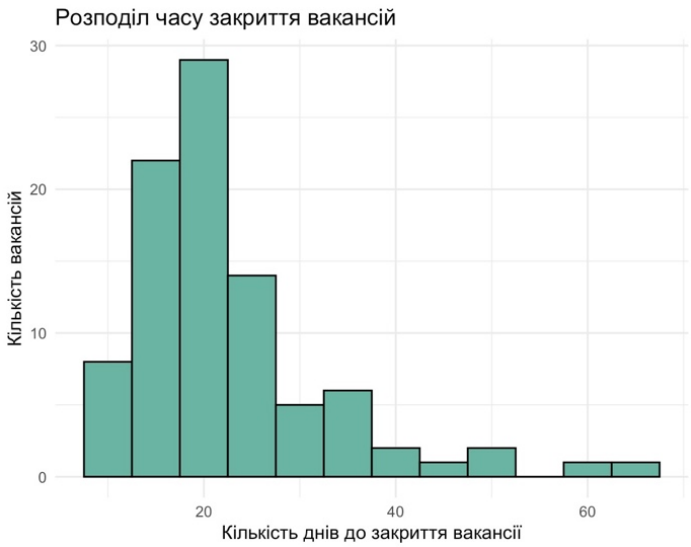

Рисунок 6 – Візуалізація розподілу ключових змінних.

Як бачимо із гістограми, розподіл є асиметричним із правостороннім зміщенням: більшість вакансій закриваються в межах 15–25 днів, тоді як окремі випадки характеризуються суттєво довшими строками (до 60–70 днів). Така структура даних свідчить про наявність вакансій із високою складністю закриття, що потенційно пов’язано зі специфічними вимогами до кандидатів або з особливостями процесу відбору.

З точки зору моделювання, це обумовлює доцільність використання алгоритмів, стійких до варіативності та викидів, а також здатних моделювати нелінійні залежності. Отримані характеристики розподілу ляжуть в основу формування аналітичної стратегії при виборі відповідної моделі прогнозування. Алгоритм Random Forest, який буде застосовано у подальшому, є стійким до нерівномірного розподілу, не вимагає нормалізації даних, толерує викиди та добре справляється з моделюванням складних нелінійних залежностей, що робить його обґрунтованим вибором для вирішення задачі прогнозування термінів закриття вакансій. Виконуємо далі кореляційний аналіз, який покаже, які змінні найбільше впливають на **days_to_close.**

Рисунок 7 – Кореляційна матриця

Найвищий рівень позитивної кореляції спостерігається між змінними *salary* та *days_to_close* (0.61), що свідчить про тенденцію до зростання тривалості процесу закриття вакансії зі збільшенням запропонованого рівня заробітної плати. Це може бути зумовлено підвищеними вимогами до кандидатів на високооплачувані посади або обмеженою кількістю фахівців відповідної кваліфікації на ринку праці. Кореляція між *num_candidates* та *days_to_close* є від'ємною (-0.29), що вказує на зворотну залежність: збільшення кількості кандидатів, як правило, сприяє скороченню тривалості рекрутингового процесу. Показник *hiring_steps* також демонструє позитивний зв'язок із *days_to_close* (0.45), що може свідчити про подовження часу закриття вакансій у разі більш складної структури етапів відбору, що потребує додаткових часових ресурсів.

Перевіримо вплив категоріальних змінних (experience_level, source, season) на days_to_close.

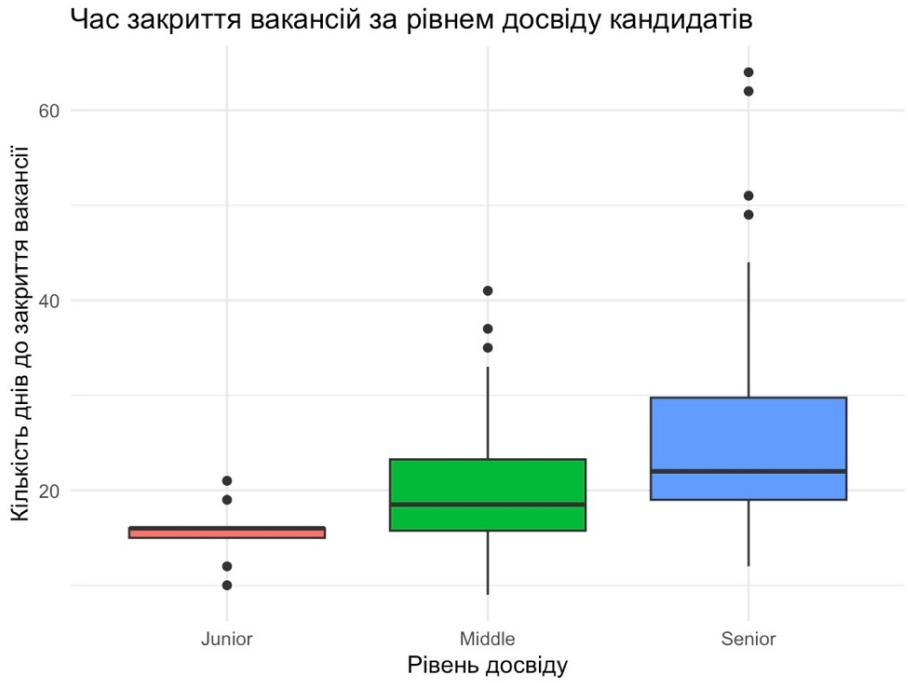

Рисунок 8 – Час закриття вакансій за рівнем досвіду.

1. Junior-позиції закриваються найшвидше

- Медіана часу закриття ≈ 15 днів (найнижча серед усіх).
- Малий розкид – більшість вакансій закриваються в діапазоні 10-20 днів.
- Майже немає викидів – стабільний процес найму.
- Ймовірні причини: висока доступність кандидатів початкового рівня та менш складні процедури відбору.

2. Middle-рівень має більший розкид

- Медіана ≈ 20 днів (довше, ніж у Junior).
- Діапазон: 10-35 днів.
- Є викиди (вакансії, які закривалися 40+ днів).
- Ймовірні причини:
    - Вимоги вищі, ніж у Junior.
    - Більше етапів відбору.
    - Конкуренція за сильних кандидатів.

3. Senior-вакансії закриваються найдовше

- Медіана ≈ 22 дні.
- Широкий розкид значень: від 10 до 60+ днів.
- Багато викидів – деякі вакансії закривалися 60+ днів.
- Ймовірні причини:
    - Складність пошуку кваліфікованих спеціалістів.
    - Вищі зарплатні очікування.
    - Довші процеси відбору (технічні інтерв'ю, тестові завдання).

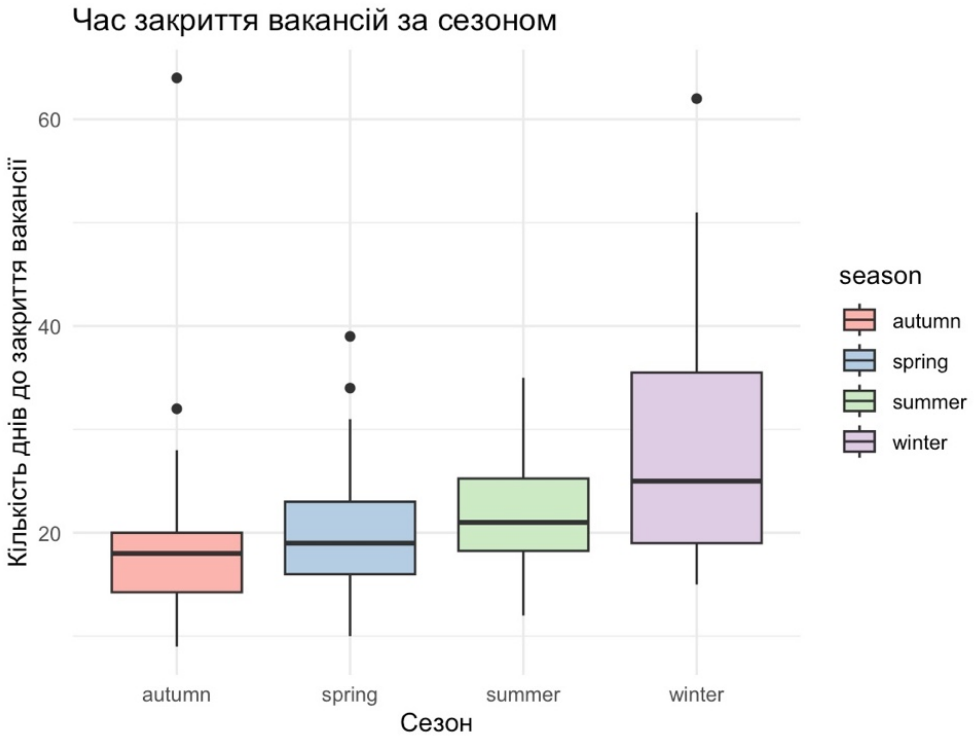

Рисунок 9 – Час закриття вакансій за сезоном.

1. Зимовий період характеризується найвищими термінами закриття вакансій.

- Медіана становить приблизно 26 днів – найвище значення серед усіх сезонів.
- Розкид значень широкий – від 10 до понад 60 днів.

- Спостерігаються численні викиди, що свідчить про нерівномірність процесу найму в цей період.

2. Вакансії, опубліковані восени, закриваються найшвидше.
   - Медіана близько 18 днів.
   - Розподіл досить щільний, із незначною кількістю викидів.
   - Більшість вакансій закриваються в межах 10-25 днів.
3. Весняний та літній періоди мають середні значення тривалості.
   - Весна – медіана приблизно 20 днів, присутні поодинокі викиди до 40 днів.
   - Літо – дещо більша варіативність, медіана на рівні 22 днів.

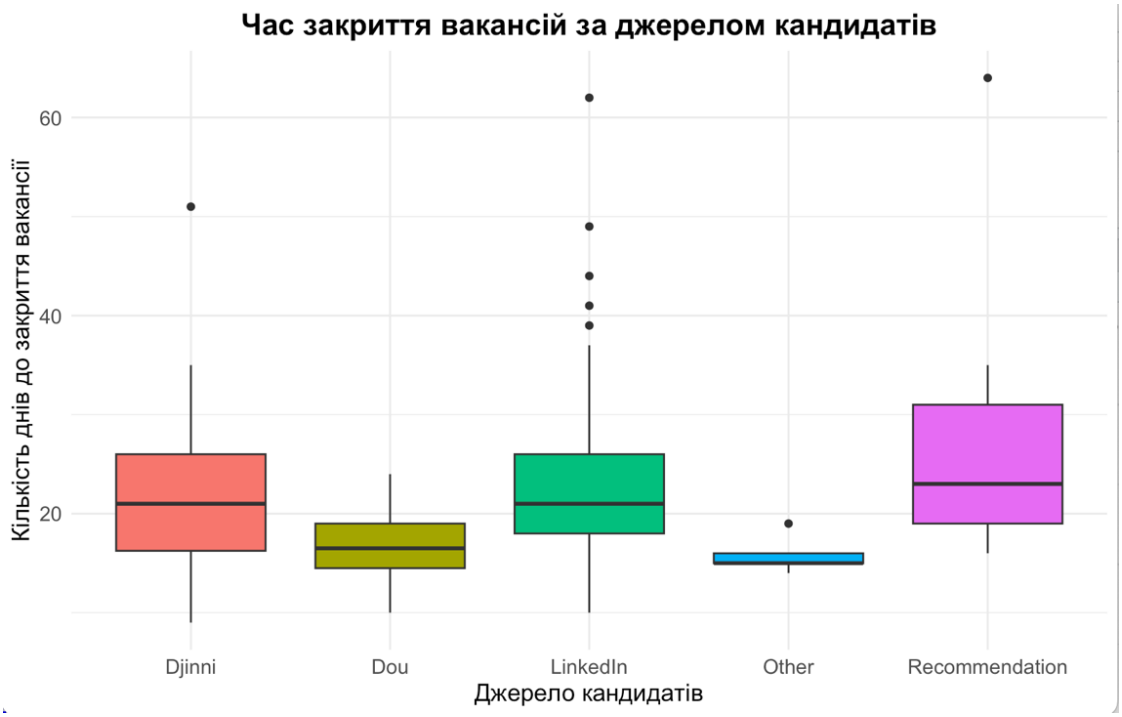

Рисунок 10 – Час закриття вакансій за джерелом кандидатів.

1. Найшвидше вакансії закриваються через джерело «Other»
   - Медіана ≈14 днів – одна з найнижчих серед усіх джерел.
   - Дуже малий розкид значень, практично відсутні викиди.
   - Це може свідчити про залучення внутрішніх кандидатів або оперативне закриття стандартних позицій.

2. «Dou» забезпечує одні з найшвидших зовнішніх наймів
    - Медіана ≈15 днів.
    - Компактний розподіл із низькою варіативністю.
    - Ймовірно, платформа використовується для оперативного пошуку доступних кандидатів.
3. «LinkedIn» та «Djinni» демонструють помірну тривалість закриття
    - Медіана становить близько 22 днів.
    - Для обох джерел характерна широка варіативність значень.
    - На «LinkedIn» виявлено низку викидів (до 60+ днів), що може свідчити про складність пошуку кваліфікованих фахівців.
4. «Recommendation» пов’язане з найдовшими строками найму
    - Медіана ≈ 26 днів.
    - Найширший розкид значень – від 10 до понад 60 днів.
    - Це джерело, ймовірно, залучається для заповнення рідкісних або складних вакансій, що потребують додаткових етапів узгодження.

Варто зазначити, що змінна job_title, попри потенційну інформативність, має надмірну деталізацію, значну кількість унікальних значень і неструктуровану класифікацію. Через це вона не використовується як предиктор у подальшому моделюванні, оскільки її обробка вимагала б окремої категоризації, що виходить за межі завдань цього дослідження. Натомість, змінна category є агрегованим представленням job_title, що дозволяє уникнути надмірної дисперсії у даних та забезпечити узагальнену класифікацію типів вакансій за напрямами діяльності (наприклад, IT, аналітика, маркетинг тощо). Завдяки обмеженій кількості категорій та логічній структурованості, ця змінна є придатною для включення до моделі як предиктор, оскільки дозволяє врахувати вплив професійного спрямування вакансії на тривалість процесу її закриття без втрати інтерпретованості та стабільності прогнозів.

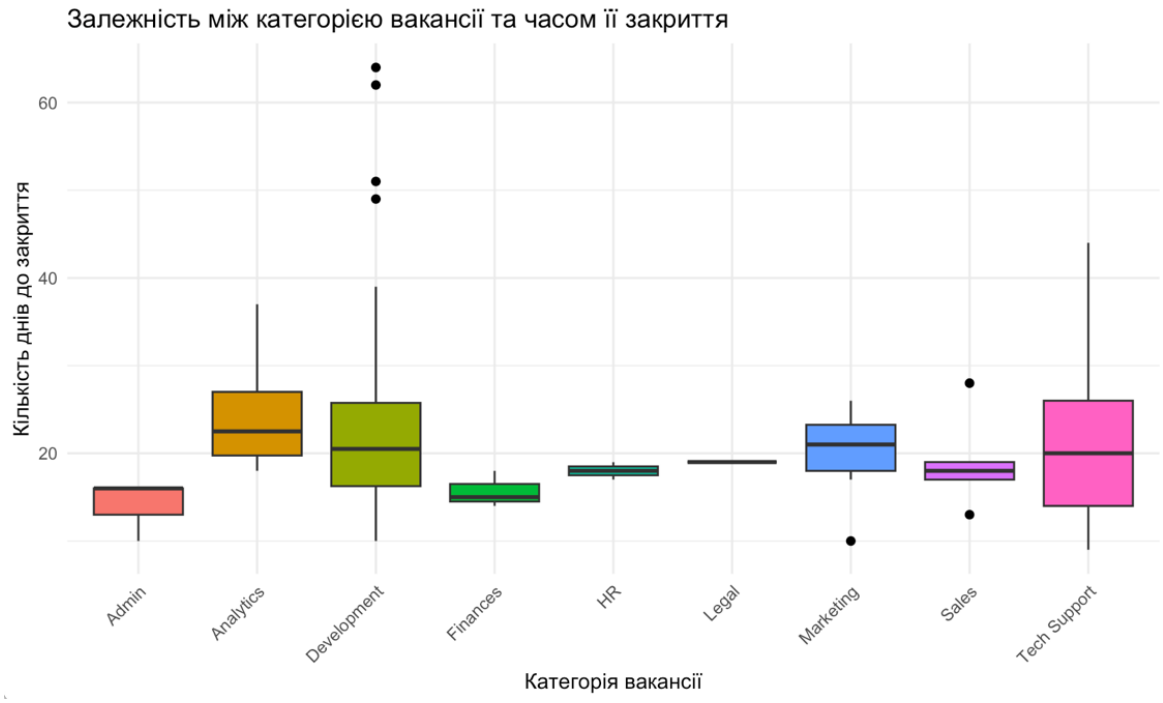

Рисунок 11 – Час закриття вакансій за категорією вакансії.

Залежність між категорією вакансії та часом її закриття демонструє помітні відмінності між професійними напрямами. Найшвидше закриваються вакансії у сфері Finances, HR і Legal – їх медіанний час становить приблизно 15 днів, а розкид значень мінімальний. Схожі темпи закриття спостерігаються й для вакансій категорії Admin, хоча розподіл дещо ширший. Вакансії категорій Analytics, Development і Tech Support характеризуються вищою медіаною (понад 20 днів) та значно більшою варіативністю. Найбільший розкид та наявність значних викидів фіксується для категорії Development – окремі вакансії закривалися понад 60 днів. Ці результати свідчать про те, що професійний напрям відіграє роль у динаміці рекрутингових процесів і може бути інформативним предиктором у подальшому моделюванні.

Незважаючи на те, що змінна salary уже була включена до кореляційного аналізу і продемонструвала найсильніший позитивний зв’язок із цільовою змінною (коефіцієнт 0.61), доцільно здійснити її додаткову візуальну перевірку. Такий аналіз дозволяє не лише підтвердити наявну залежність, а й дослідити її форму – лінійну або нелінійну, а також виявити можливі викиди чи зони із

підвищеною концентрацією значень. Це забезпечує глибше розуміння характеру впливу рівня заробітної плати на тривалість закриття вакансій і може бути корисним при виборі моделі прогнозування.

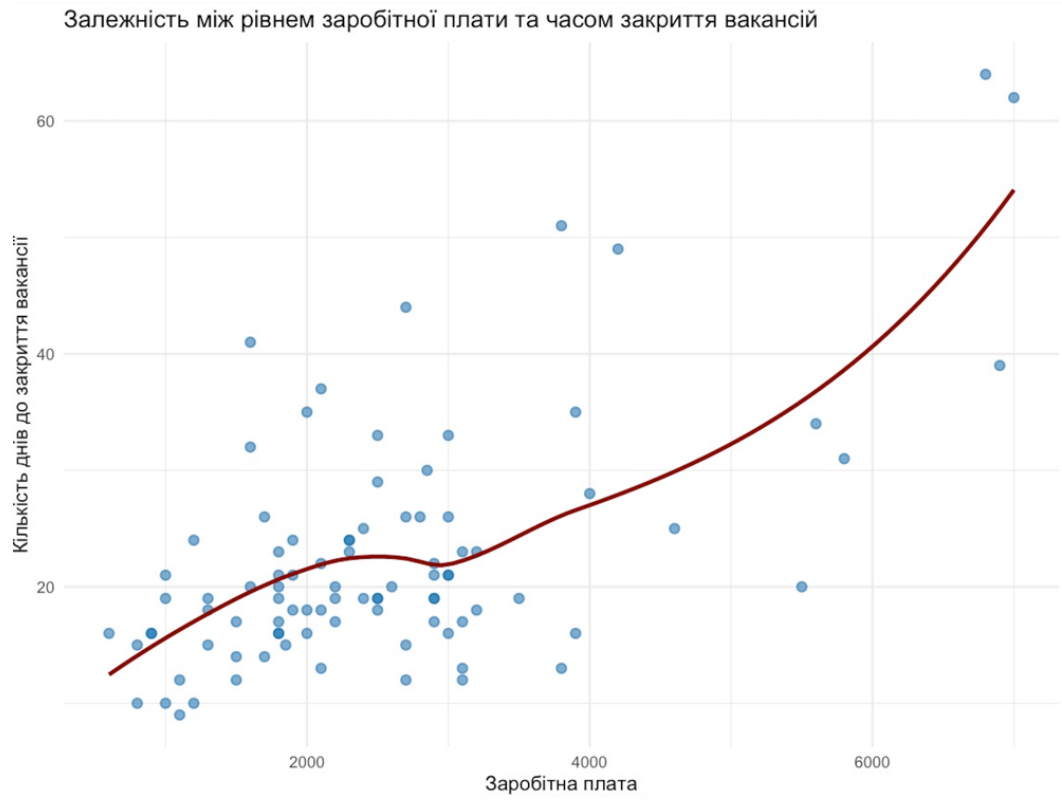

Рисунок 12 – Залежність між рівнем заробітної плати та часом закриття вакансій.

Графік розсіювання з накладеною трендовою кривою демонструє чітко виражену позитивну нелінійну залежність між рівнем заробітної плати та тривалістю закриття вакансії. Основна маса вакансій із заробітною платою до 3000 умовних одиниць закривається у діапазоні 15-30днів. Починаючи з позначки близько 4000, спостерігається зростання кількості випадків із тривалим періодом найму – до 60 і більше днів. Це підтверджує гіпотезу про те, що високооплачувані вакансії потребують більше часу на закриття, ймовірно, через складніші вимоги до кандидатів або обмежену пропозицію фахівців відповідного рівня. Такий розподіл підкреслює доцільність застосування методу Random Forest, який здатен враховувати складні нелінійні взаємозв’язки між змінними.

Проведений попередній аналіз приводить до таких ключових висновків:

1. Кореляційний аналіз показав, що змінні salary, hiring_steps та num_candidates мають найвищий рівень зв’язку з days_to_close: підвищення зарплати та кількості етапів відбору асоціюється з довшим терміном закриття вакансій, тоді як більша кількість кандидатів, навпаки, сприяє його скороченню.
2. Рівень досвіду кандидата впливає на швидкість найму: junior-позиції закриваються найшвидше, а senior-позиції – найдовше, що свідчить про складність пошуку кваліфікованих спеціалістів.
3. Джерело надходження кандидатів демонструє різну ефективність: найшвидші закриття вакансій відбуваються через канали Dou та Other, тоді як Recommendation характеризується найдовшими термінами.
4. Сезонний фактор також має вплив: вакансії, опубліковані восени, закриваються швидше, тоді як у зимовий період спостерігається максимальна тривалість рекрутингового процесу.
5. Категорія вакансії є важливою змінною – вакансії у сфері Finances, HR та Legal закриваються значно швидше, ніж у категорії Development, яка демонструє найбільшу варіативність.
6. Візуальний аналіз залежності між salary та days_to_close підтверджує наявність нелінійного зв’язку: високооплачувані вакансії характеризуються більшою тривалістю закриття, що обґрунтовує вибір моделі, здатної враховувати складні взаємозв’язки – зокрема Random Forest.

Варто також додати, що структура датасету є достатньо збалансованою для побудови прогнозної моделі, більшість змінних мають коректні типи та відсутні пропущені значення. Виявлені статистичні характеристики та закономірності дозволяють визначити найбільш інформативні предиктори для прогнозування цільової змінної. Отримані результати лягають в основу подальшого етапу –

побудови та налаштування моделі машинного навчання, яка дозволить реалізувати задачу прогнозування на практиці.

3.2 Побудова моделі методом Random Forest

З урахуванням результатів попереднього аналітичного етапу для побудови прогнозної моделі обрано метод Random Forest, який поєднує високу точність, стійкість до перенавчання та здатність обробляти як числові, так і категоріальні ознаки без необхідності попередньої трансформації. На цьому етапі здійснюється безпосереднє навчання моделі, що включає формування навчальної та тестової вибірок, налаштування гіперпараметрів, а також побудову моделі на основі навчальних даних для подальшої оцінки її ефективності.

Як вже зазначалось, для навчання моделі використано синтетично згенерований датасет, який максимально наближений до реальних даних та базується на рекрутингових даних, зібраних під час проходження науково-дослідної практики із використанням системи HRIS BambooHR. Цей набір охоплює ключові характеристики вакансій, процесу найму та джерел кандидатів, зберігаючи логіку та статистичні властивості, притаманні реальним даним. Завантаження даних до середовища R здійснювалося з Excel-файлу за допомогою функції read_excel() з пакету readxl.

В рамках попередньої обробки даних було здійснено перетворення категоріальних змінних – experience_level, category, source, season – у фактори для забезпечення їхньої коректної інтерпретації під час побудови моделі. Необхідним етапом є також видалення тих змінних, що не мають прогностичної цінності або не використовуються безпосередньо у моделюванні. Зокрема, змінна job_title не є інформативною для побудови моделі та може бути виключена з датасету з метою запобігання надмірній складності та потенційній кореляційній зашумленості. Крім того, для уникнення витоку інформації було вилучено змінні open_date та close_date, на основі яких розраховується цільова змінна days_to_close. Це дозволяє гарантувати, що модель здійснює

прогнозування виключно на основі доступних на момент відкриття вакансії даних, а не на основі похідних часових характеристик. З тієї ж причини було виключено змінну num_candidates, оскільки інформація про кількість кандидатів стає відомою лише постфактум і не може використовуватися для передбачення на етапі відкриття вакансії.

```
> data <- data %>%
+     mutate(days_to_close = as.numeric(difftime(close_date, open_date, units = "days"))) %
>%
+     select(-open_date, -close_date, -num_candidates, -job_title)
> str(data)
tibble [91 × 7] (S3: tbl_df/tbl/data.frame)
 $ category        : chr [1:91] "Analytics" "Development" "Development" "Tech Support" ...
 $ experience_level: chr [1:91] "Middle" "Senior" "Junior" "Senior" ...
 $ salary          : num [1:91] 2100 3900 1000 2700 1600 7000 2300 2850 4200 1600 ...
 $ days_to_close   : num [1:91] 37 35 21 44 41 62 24 30 49 20 ...
 $ hiring_steps    : num [1:91] 3 4 4 3 3 4 4 3 4 3 ...
 $ source          : chr [1:91] "LinkedIn" "Recommendation" "Dou" "LinkedIn" ...
 $ season          : chr [1:91] "winter" "winter" "winter" "winter" ...
> summary(data$days_to_close)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   9.00   16.00   19.00   22.46   25.00   64.00
```

Рисунок 13 – Код для видалення змінних.

Переконаємось, що змінні успішно видалені:

```
> names(data)
[1] "category"         "experience_level" "salary"           "days_to_close"
[5] "hiring_steps"     "source"           "season"
```

Рисунок 14 – Застосування функції names.

Бачимо, що змінні job_title та num_candidates успішно видалена – їх немає серед назв стовпців у data. Крім того, з датасету вилучено змінні open_date та close_date, які використовувались лише для обчислення цільової змінної. Натомість змінна days_to_close була створена шляхом розрахунку різниці між датами відкриття та закриття вакансії, і наразі містить числові значення, що відповідають кількості днів до закриття. Таким чином, набір даних повністю підготовлений до моделювання.

Оскільки попередню обробку даних завершено, всі змінні приведено до відповідних типів, а непотрібні стовпці видалено, наступним етапом є підготовка вибірки для побудови прогнозної моделі. З метою забезпечення об'єктивної оцінки якості моделі, датасет буде розділено на навчальну (train) та тестову (test) підвибірки, що дозволить здійснити тренування моделі на одному наборі даних і перевірити її узагальнювальну здатність на іншому. Навчальна вибірка буде використана для побудови моделі Random Forest, тоді як тестова – для перевірки точності прогнозування на нових, раніше не бачених даних. Поділ здійснюється у співвідношенні 80:20, що є загальноприйнятою практикою в задачах машинного навчання. Для забезпечення відтворюваності результатів фіксується початкове значення генератора випадкових чисел.

```
> set.seed(123)
> train_indices <- sample(seq_len(nrow(data)), size = 0.8 * nrow(data))
>
> train_data <- data[train_indices, ]
> test_data <- data[-train_indices, ]
```

Рисунок 15 – Створення вибірок.

Після остаточної підготовки даних та формування навчальної і тестової вибірок можна переходити до етапу навчання моделі Random Forest, яка була обґрунтована як основний алгоритм прогнозування термінів закриття вакансій у попередніх розділах.

```
> set.seed(123)
> rf_model <- randomForest(
+     days_to_close ~ .,
+     data = train_data,
+     ntree = 500,
+     importance = TRUE
+ )
> print(rf_model)
```

Рисунок 16 – застосування методу Random Forest.

Для навчання моделі було використано 500 дерев (ntree = 500), а кількість змінних, які розглядаються при кожному розщепленні дерева (mtry), залишено на рівні значення за замовчуванням, що відповідає загальноприйнятому підходу

для задач регресії – p / 3, де p позначає кількість змінних-предикторів. У цьому випадку за наявності семи змінних, що використовуються для передбачення, параметр mtry набуває значення 2. Додатково параметр importance = TRUE дозволяє оцінити внесок кожної змінної у формування прогнозу. Після навчання було здійснено виведення характеристик побудованої моделі з використанням функції print(), результати якої наведено нижче.

```
Call:
 randomForest(formula = days_to_close ~ ., data = train_data,      ntree = 500, importance =
TRUE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 2

          Mean of squared residuals: 67.97641
                    % Var explained: 38.84
```

Рисунок 17 – Виведення результатів моделі.

Побудована модель Random Forest наразі пояснює 38.84% варіації цільової змінної days_to_close, що свідчить про обмежену, але наявну здатність моделі виявляти залежності в даних. Для перевірки точності прогнозування протестуємо модель на тестовій підвибірці, яка включає 19 вакансій, не використаних у процесі навчання. Як метрики оцінювання застосуємо середню абсолютну похибку (MAE) та корінь середньоквадратичної похибки (RMSE), що дозволяє кількісно оцінити середнє відхилення прогнозованих значень від фактичних у днях.

```
> predictions <- predict(rf_model, newdata = test_data)
> mae_result <- mean(abs(test_data$days_to_close - predictions))
> rmse_result <- sqrt(mean((test_data$days_to_close - predictions)^2))
> cat("MAE:", round(mae_result, 2), "\n")
MAE: 5.13
> cat("RMSE:", round(rmse_result, 2), "\n")
RMSE: 7.18
```

Рисунок 18 – Оцінка точності моделі.

Отримані значення метрик RMSE (Root Mean Square Error) та MAE (Mean Absolute Error) становлять відповідно 7.18 та 5.13 Це означає, що середнє відхилення прогнозованих значень терміну закриття вакансій від фактичних становить приблизно 5 днів, а середньоквадратичне – понад 7 днів. Зважаючи на те, що в умовах практичного рекрутингу середня тривалість процесу закриття вакансії зазвичай не перевищує 30 днів, така точність є прийнятною лише для задач попереднього планування. Водночас у ситуаціях, де своєчасність найму критично впливає на бізнес-процеси або строки реалізації проєктів, зазначений рівень похибки може бути недостатнім. Отже, є підстави для подальшого вдосконалення моделі з метою підвищення її прогностичної здатності.

З метою глибшого розуміння структури побудованої моделі та визначення потенційних напрямів її вдосконалення здійснимо аналіз важливості змінних за допомогою функції importance(). Це дозволить оцінити, які саме фактори мають найбільший вплив на прогнозування тривалості закриття вакансій у межах побудованої моделі.

```
> importance(rf_model)
                  %IncMSE IncNodePurity
category         2.272855      583.7051
experience_level 4.609336      484.0700
salary          17.224781     3311.3414
hiring_steps     3.149099      598.6563
source           5.823895      636.7201
season           8.255697      968.9541
```

Рисунок 19 – Аналіз важливості змінних моделі.

На основі результатів аналізу важливості змінних отримали дві метрики:

- **%IncMSE (Mean Decrease in Accuracy)** – наскільки сильно зростає помилка, якщо прибрати змінну.
- **IncNodePurity (Total Decrease in Node Impurity)** – наскільки змінна допомагає ділити дані у вузлах дерева.

Встановлено, що найбільший внесок у зниження похибки прогнозування має змінна salary (%IncMSE = 17.22), що свідчить про суттєвий вплив рівня заробітної плати на тривалість закриття вакансії. Також виразну прогностичну

силу демонструє змінна season (%IncMSE = 8.26), що може відображати вплив сезонних коливань активності на ринку праці. Серед змінних із помірною важливістю варто відзначити source (5.82%), experience_level (4.61%) та hiring_steps (3.15%), які пов’язані з каналом пошуку, профілем кандидата та складністю етапів добору відповідно. Змінна category продемонструвала найнижчий рівень впливу (%IncMSE = 2.27), що може вказувати на її обмежену роль у поясненні варіативності цільової змінної в межах цієї моделі.

З огляду на помірний рівень точності базової моделі Random Forest, а також на результати аналізу важливості предикторів, доцільно здійснити подальше налаштування моделі з метою підвищення її прогностичної ефективності. У цьому контексті наступним етапом є реалізація гіперпараметричної оптимізації, що передбачає підбір оптимальних поєднань параметрів моделі на основі результатів крос-валідаційного тестування.

З метою підвищення прогностичної здатності моделі буде здійснено крос-валідаційну оптимізацію з використанням сіткового пошуку за декількома ключовими гіперпараметрами моделі Random Forest, зокрема mtry (кількість змінних, що розглядаються на кожному розщепленні дерева), nodesize (мінімальна кількість спостережень у листі) та maxnodes (максимальна кількість вузлів у дереві). Для цього буде використано 5-кратну крос-валідацію, що дозволяє забезпечити надійну оцінку якості моделі на різних підвибірках та мінімізувати ризик переобучення. Оптимальна комбінація гіперпараметрів визначатиметься на основі мінімального значення метрики RMSE, що характеризує середнє квадратичне відхилення прогнозованих значень від фактичних.

Для забезпечення можливості одночасної оптимізації декількох гіперпараметрів моделі Random Forest, зокрема mtry, nodesize та maxnodes, створимо спеціалізований метод customRF, який дозволяє реалізувати налаштування цих параметрів у межах фреймворку caret. Оскільки базова реалізація методу "rf" у пакеті caret не підтримує безпосередню оптимізацію параметрів nodesize та maxnodes, виникає потреба у створенні користувацької

обгортки, яка визначає порядок тренування моделі та спосіб передачі гіперпараметрів.

```
> set.seed(123)
> tune_grid <- expand.grid(
+     mtry = c(2, 3, 4, 5),
+     nodesize = c(5, 10, 15),
+     maxnodes = c(10, 20, 30)
+ )
> customRF <- list(
+     type = "Regression",
+     library = "randomForest",
+     loop = NULL,
+     parameters = data.frame(
+         parameter = c("mtry", "nodesize", "maxnodes"),
+         class = rep("numeric", 3),
+         label = c("mtry", "nodesize", "maxnodes")
+     ),
+     grid = function(x, y, len = NULL, search = "grid") {},
+     fit = function(x, y, wts, param, lev, last, classProbs, ...) {
+         randomForest(
+             x, y,
+             mtry = param$mtry,
+             nodesize = param$nodesize,
+             maxnodes = param$maxnodes,
+             ntree = 500,
+             importance = TRUE, ...
+         )
+     },
+     predict = function(modelFit, newdata, submodels = NULL) {
+         predict(modelFit, newdata)
+     },
+     prob = NULL,
+     sort = function(x) x[order(x[,1]),],
+     levels = function(x) {}
+ )
```

Рисунок 20 – Гіперпараметризація моделі.

Об’єкт customRF задає ключові компоненти, необхідні для інтеграції з функцією train(): тип задачі (Regression), виклик базової бібліотеки (randomForest), перелік параметрів, які можуть бути оптимізовані, а також функції для навчання (fit) та прогнозування (predict). Таким чином, customRF формалізує процедуру побудови моделей з нетиповими налаштуваннями дерева рішень і дозволяє виконати сітковий пошук за заданими комбінаціями гіперпараметрів у процесі крос-валідаційного тестування. На основі створеного об'єкта customRF здійснюється навчання моделі з використанням функції train(),

де за допомогою аргументів x та y передаються відповідно матриця предикторів і вектор цільових значень. Контроль процесу навчання реалізується через параметр trControl, в якому задається схема 5-кратної крос-валідації. Аргумент tuneGrid містить сітку значень для трьох гіперпараметрів – mtry, nodesize та maxnodes, які буде послідовно протестовано для виявлення оптимальної конфігурації моделі. Таким чином, здійснюється повнофакторний перебір комбінацій параметрів із подальшим вибором тієї, що забезпечує найнижчий рівень помилки прогнозування за метрикою RMSE.

```
> ctrl <- trainControl(method = "cv", number = 5)
> tuned_rf <- train(
+     x = subset(train_data, select = -days_to_close),
+     y = train_data$days_to_close,
+     method = customRF,
+     trControl = ctrl,
+     tuneGrid = tune_grid
+ )
There were 50 or more warnings (use warnings() to see the first 50)
> print(tuned_rf$bestTune)
   mtry nodesize maxnodes
36    5       15       30
> optimized_rf <- randomForest(
+     days_to_close ~ .,
+     data = train_data,
+     ntree = 500,
+     mtry = tuned_rf$bestTune$mtry,
+     nodesize = tuned_rf$bestTune$nodesize,
+     maxnodes = tuned_rf$bestTune$maxnodes,
+     importance = TRUE
+ )
```

Рисунок 20 – Навчання оптимізованої моделі.

Після побудови остаточної моделі Random Forest з використанням оптимальної комбінації гіперпараметрів, визначеної в результаті крос-валідаційного сіткового пошуку, необхідно здійснити її оцінку на тестовій вибірці. Це дозволить перевірити, наскільки добре модель узагальнює закономірності та здатна здійснювати точне прогнозування для нових, раніше не бачених даних.

```
> pred_optimized <- predict(optimized_rf, newdata = test_data)
>
>
> actuals <- test_data$days_to_close
>
> rmse_final <- sqrt(mean((pred_optimized - actuals)^2))
> mae_final  <- mean(abs(pred_optimized - actuals))
>
> cat("RMSE (оптимізована модель):", round(rmse_final, 2), "\n")
RMSE (оптимізована модель): 8.04
> cat("MAE (оптимізована модель):", round(mae_final, 2), "\n")
MAE (оптимізована модель): 5.79
```

Рисунок 21 - Оцінка точності оптимізованої моделі.

Незважаючи на застосування гіперпараметричної оптимізації за параметрами mtry, nodesize та maxnodes, отримана модель не продемонструвала загального покращення точності порівняно з базовою. Зокрема, для базової моделі значення MAE становить 5.13, а RMSE – 7.18, тоді як для оптимізованої моделі ці показники становлять відповідно 5.79 і 8.04. Незважаючи на спробу вдосконалення моделі шляхом налаштування гіперпараметрів, вона не продемонструвала покращення результатів, а навпаки – збільшення середньоквадратичної похибки свідчить про підвищену варіативність помилок. Це може бути зумовлено перенавчанням або надмірною складністю моделі. Отже, у межах наявного обсягу даних доцільно зберегти параметри базової моделі, які забезпечують більш стабільні результати прогнозування.

3.3. Оцінка якості моделі

Оцінка якості побудованої моделі є ключовим етапом у процесі аналізу, оскільки дозволяє встановити, наскільки ефективно модель виконує своє основне завдання – прогнозування термінів закриття вакансій. З метою перевірки ефективності прогнозування термінів закриття вакансій методом Random Forest у цьому дослідженні використано дві ключові метрики – MAE (середня абсолютна похибка) та RMSE (корінь з середньої квадратичної похибки). Ці

метрики вимірюють точність прогнозів у тих самих одиницях, що й цільова змінна – у днях, – що забезпечує інтуїтивну інтерпретацію результатів.

Таблиця 3.3.1. Порівняння метрик якості моделей

Модель	**MAE**	**RMSE**
Random Forest (базова)	5.13	7.18
Random Forest (оптимізована)	5.79	8.04

Отримані результати свідчать, що базова модель продемонструвала загалом кращу точність, ніж оптимізована конфігурація. Значення MAE у базової моделі становить 5.13, а в оптимізованої – 5.79, що вказує на меншу середню абсолютну похибку у базовому варіанті. Крім того, RMSE у базової моделі також нижчий (7.18 проти 8.04), що свідчить про меншу варіативність похибок. Це може вказувати на зниження здатності оптимізованої моделі до узагальнення через ускладнення її структури. У середньому базова модель допускала похибку у межах 5-7 днів, що є прийнятним рівнем точності для попереднього планування термінів найму.

Для якісного аналізу результатів побудуємо графік розсіювання прогнозованих проти фактичних значень.

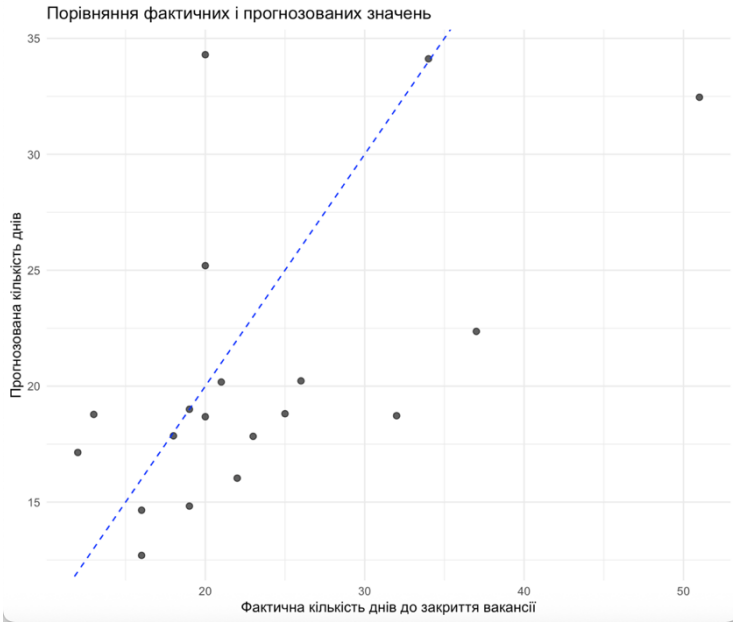

Рисунок 21 – Графік розсіювання прогнозованих проти фактичних значень.

Графік розсіювання свідчить про помірну відповідність між фактичними та прогнозованими значеннями. Більшість точок розташована поблизу діагоналі, що вказує на прийнятну точність у межах основного діапазону спостережень. Водночас у граничних значеннях помітні відхилення, переважно у напрямку переоцінки тривалості закриття вакансії, коли модель прогнозує довший термін, ніж відбулося насправді.

У контексті HR-планування та управління ресурсами така похибка є менш критичною, ніж недооцінка. Завищений прогноз дозволяє закласти додатковий час на підбір персоналу, уникнути завищених очікувань у менеджменту та краще розподілити навантаження між рекрутерами. Натомість недооцінка могла б призвести до нереалістичних термінів закриття позицій, порушення графіків проєктів або втрати довіри до HR-аналітики. Таким чином, характер похибок моделі є прийнятним для практичного застосування у внутрішньому плануванні.

Підсумовуючи результати проведеного моделювання та оцінки якості прогнозів, можна зробити висновок про доцільність практичного застосування побудованої моделі у сфері управління персоналом. Незважаючи на певні обмеження, модель на основі Random Forest продемонструвала прийнятний рівень точності, що дозволяє використовувати її для формування орієнтовних оцінок термінів закриття вакансій. Зокрема, у ситуаціях, коли організація не має у своєму розпорядженні великого обсягу історичних даних або доступ до змінних є обмеженим, запропонований підхід може відігравати роль ефективного допоміжного інструменту. Модель може бути інтегрована у внутрішні аналітичні процеси для підтримки планування рекрутингових кампаній, встановлення реалістичних термінів підбору персоналу та управління очікуваннями стейкхолдерів. Таким чином, навіть за умов певної варіативності прогнозів, модель може сприяти підвищенню обґрунтованості управлінських рішень у HR-практиці.

3.4 Перспективи подальших досліджень

Розроблена в межах цього дослідження модель прогнозування термінів закриття вакансій, побудована на основі алгоритму Random Forest, продемонструвала задовільний рівень точності та практичну придатність у задачах попереднього HR-планування. Проте, з огляду на низку концептуальних і прикладних обмежень, існують вагомі підстави для подальшого вдосконалення як самої моделі, так і методологічного підходу загалом. Перспективні напрями розвитку можна окреслити за кількома взаємопов'язаними аспектами.

Один із напрямів удосконалення моделі полягає в розширенні набору вхідних змінних. У даному дослідженні використовувалися лише загальні параметри вакансії, проте не враховувалися деякі важливі характеристики, що можуть мати суттєвий вплив на тривалість найму. Серед таких факторів доцільно розглядати тип працевлаштування (постійна зайнятість, контракт, фріланс), вимоги до ключових компетенцій, рівень конкуренції на ринку (кількість схожих вакансій), популярність посади в конкретному регіоні, наявність внутрішньої заміни тощо. Інтеграція таких змінних дозволить створити більш повну картину умов найму і, відповідно, підвищити точність прогнозу.

Ще одним важливим аспектом є вивчення альтернативних підходів до моделювання. Хоча алгоритм Random Forest продемонстрував ефективність у роботі з табличними даними, інші підходи можуть забезпечити кращу якість прогнозування за рахунок гнучкішої структури або вбудованої оптимізації. Зокрема, доцільно дослідити можливості застосування градієнтного бустингу, зокрема алгоритму XGBoost, який часто демонструє переваги у задачах регресії за рахунок високої точності, роботи з пропущеними значеннями та ефективного контролю перенавчання. Крім того, варто розглянути комбінування моделей: наприклад, застосування простої лінійної регресії для базового прогнозування у типових ситуаціях і складніших моделей – для обробки нетипових або критичних вакансій.

Також актуальним є питання інтеграції моделі у бізнес-процеси компанії. У перспективі модель може бути імплементована у складі корпоративних систем управління персоналом (HRIS) або систем підтримки прийняття рішень. Такий підхід дозволить автоматизувати процес оцінювання очікуваних термінів закриття вакансій, що, у свою чергу, сприятиме підвищенню ефективності планування, кращому розподілу навантаження між рекрутерами, а також зменшенню ризику зриву проєктних термінів через кадрові затримки.

Окремої уваги заслуговує можливість реалізації механізмів адаптивного навчання (online learning), коли модель регулярно оновлюється на основі нових даних щодо вакансій та реальних строків їх закриття. Такий підхід дозволить забезпечити динамічну адаптацію моделі до змін кон'юнктури ринку праці, сезонних коливань активності кандидатів та інших часових факторів. Застосування регулярного переобучення може значно підвищити релевантність прогнозів у довгостроковій перспективі.

Загалом, подальші дослідження у цьому напрямі доцільно орієнтувати на підвищення прогностичної здатності моделей, зростання рівня інтерпретованості результатів для бізнес-користувача та їх інтеграцію у реальні управлінські практики HR-аналітики.

Висновки

У третьому розділі здійснено практичну реалізацію методології прогнозування термінів закриття вакансій із використанням технологій Data Science. Було розроблено та апробовано повноцінну аналітичну модель, яка охоплює усі ключові етапи побудови прогнозного рішення: від попередньої обробки даних і виявлення структурних закономірностей до навчання моделі машинного навчання, оцінки її ефективності та пошуку можливостей для подальшого вдосконалення. Результати, отримані в межах даного розділу, дозволяють сформулювати низку важливих висновків, що підтверджують доцільність і практичну релевантність обраного підходу.

Передусім було підтверджено якість синтетично згенерованого датасету, який відповідає структурі й логіці реальних рекрутингових даних. Проведений описовий аналіз дозволив виявити основні закономірності, що впливають на тривалість найму: позитивний зв’язок між рівнем заробітної плати, кількістю етапів відбору та часом закриття вакансій, а також зворотну залежність від кількості кандидатів. Аналіз категоріальних змінних засвідчив відмінності у тривалості рекрутингових процесів залежно від рівня досвіду кандидатів, джерел пошуку, сезону публікації вакансії та її професійної категорії.

На основі виявлених закономірностей було здійснено побудову моделі методом Random Forest, що забезпечує стійкість до викидів, здатність враховувати нелінійні зв’язки та підтримку роботи з факторними змінними без потреби у додатковому кодуванні. Базова модель продемонструвала прийнятні результати за метриками точності, що дозволяє застосовувати її для попереднього планування строків найму. Застосування гіперпараметричної оптимізації не призвело до суттєвого покращення якості моделі, що свідчить про доцільність використання спрощеної конфігурації в умовах обмеженого обсягу даних. Аналіз важливості предикторів виявив, що ключовими змінними, які визначають час закриття вакансії, є рівень заробітної плати та сезон найму. Також помірну прогностичну силу мають джерело пошуку, рівень досвіду кандидата та складність процесу відбору. Це підтверджує практичну цінність отриманих висновків для HR-аналітики, оскільки дозволяє зосередити увагу на оптимізації саме тих факторів, які мають найбільший вплив на ефективність рекрутингового процесу.

Таким чином, результати, отримані у межах третього розділу, підтверджують доцільність використання методу Random Forest для прогнозування термінів закриття вакансій. Побудована модель здатна підтримувати процес прийняття управлінських рішень у сфері планування персоналу. Водночас результати моделювання окреслюють перспективи подальших досліджень, зокрема щодо розширення набору ознак, тестування альтернативних алгоритмів та інтеграції моделі у реальні бізнес-процеси.

РОЗДІЛ 4
ТЕХНОЛОГІЯ ВПРОВАДЖЕННЯ МОДЕЛІ ПРОГНОЗУВАННЯ ТЕРМІНІВ ЗАКРИТТЯ ВАКАНСІЙ У ПРАКТИЧНУ ДІЯЛЬНІСТЬ

4.1 Формування концепції застосування прогнозної моделі у процесах управління персоналом

В умовах цифровізації управлінських процесів та підвищення вимог до ефективності HR-відділів все більшої актуальності набувають інструменти, які забезпечують підтримку прийняття обґрунтованих управлінських рішень на основі аналізу даних. Одним із таких інструментів є система прогнозування термінів закриття вакансій, яка дозволяє підвищити точність планування кадрових ресурсів і зменшити ризики, пов’язані з неочікуваними затримками у рекрутингових процесах.

У цьому контексті модель прогнозування термінів закриття вакансій може бути розглянута як складова частина системи підтримки прийняття рішень (Decision Support System, DSS) у сфері управління персоналом. DSS – це концептуальна та технологічна система, яка забезпечує керівників і фахівців можливістю ухвалювати обґрунтовані рішення на основі обробки й аналізу релевантних даних. Основна мета DSS полягає не у повній автоматизації процесів, а в зменшенні рівня невизначеності та підвищенні якості управлінських рішень через надання інформативних прогнозів, оцінок та сценаріїв.

Класична структура DSS включає три основні компоненти:

- база даних, яка забезпечує зберігання та доступ до історичної інформації;
- база моделей, яка виконує аналітичні обчислення або прогнозування;
- інтерфейс користувача, що забезпечує взаємодію з системою та візуалізацію результатів [8].

У випадку запропонованої моделі ці компоненти реалізуються таким чином:

- дані формуються на основі історичних вакансій, які містять ключові характеристики позицій та тривалість їх закриття;
- модель Random Forest виступає аналітичним ядром системи, що забезпечує прогноз тривалості найму;
- інтерфейс може бути реалізований у вигляді форми введення параметрів нової вакансії та отримання прогнозу (наприклад, у Google Sheets, Excel або HRIS).

У межах даного дослідження модель Random Forest розглядається як ядро DSS, яке може бути інтегроване в існуючі HR-процеси компанії без необхідності впровадження окремої спеціалізованої платформи. Система може функціонувати у вигляді легкої інтеграції на базі вже наявних інструментів, зокрема R, Excel або Power BI, що робить її доступною навіть для організацій із обмеженими ресурсами.

Важливим елементом концепції впровадження є чітке розуміння ролей користувачів, які взаємодіятимуть із моделлю в межах організаційного процесу. Основними категоріями користувачів є:

- HR-аналітик – відповідає за технічне налаштування та запуск моделі, оновлення даних і контроль за точністю прогнозів;
- рекрутер – використовує модель для планування своєї роботи, визначення послідовності обробки заявок та узгодження термінів із замовниками;
- менеджер напряму або керівник підрозділу – отримує зведену інформацію про прогнозні строки закриття вакансій для прийняття управлінських рішень щодо ресурсів, пріоритетів чи організації процесів;
- керівництво HR-департаменту – використовує аналітичні звіти для оцінки ефективності підбору, виявлення «вузьких місць» у рекрутингу та обґрунтування змін у стратегії залучення персоналу.

Отже, розроблена модель має потенціал для інкорпорації у щоденну практику управління персоналом, забезпечуючи підтримку як у короткостроковому плануванні дій з підбору персоналу, так і у вирішенні питань довгострокового розподілу та прогнозування потреб у кадрах.

Призначення розробленої прогнозної моделі у сфері управління персоналом полягає насамперед у підвищенні обґрунтованості та точності рекрутингового планування. У практичному вимірі це означає, що HR-фахівці можуть не лише покладатися на досвід або інтуїцію при оцінці тривалості закриття вакансій, а й використовувати формалізовану модель, яка ґрунтується на аналізі історичних даних. Модель надає кількісну оцінку очікуваної тривалості найму залежно від характеристик вакансії. Це дозволяє ефективніше координувати дії всіх учасників процесу – від рекрутерів до керівників функціональних підрозділів.

Зокрема, прогноз може використовуватись для:

- визначення пріоритетності вакансій;
- планування черговості заповнення позицій;
- встановлення реалістичних дедлайнів;
- оцінки ризику затягування найму для кожного окремого запиту.

Основними цілями впровадження моделі є:

- підвищення точності планування рекрутингових кампаній за рахунок використання історичних даних і прогнозних алгоритмів;
- оптимізація розподілу робочого навантаження між рекрутерами з урахуванням складності вакансій;
- забезпечення прозорості у взаємодії між HR-відділом та керівниками підрозділів завдяки формуванню реалістичних очікувань щодо строків підбору персоналу;
- підтримка прийняття управлінських рішень у ситуаціях обмежених ресурсів;
- формування персоналізованих показників ефективності (KPI) для рекрутерів.

Застосування моделі як інструменту підтримки планування дозволяє зменшити кількість суб’єктивних рішень і сприяє переходу до практики управління, заснованої на даних. Крім того, модель відповідає принципам адаптивності та масштабованості, що є характерними для сучасних DSS. У разі

накопичення нових даних її можна регулярно перенавчати, враховуючи зміни ринку праці, поведінки кандидатів і внутрішні трансформації бізнес-процесів.

Таким чином, концепція впровадження моделі полягає у створенні гнучкого інструменту підтримки прийняття рішень, що інтегрується в рекрутинговий цикл і дозволяє підвищити ефективність управління людськими ресурсами шляхом обґрунтованого прогнозування часових параметрів найму.

4.2 Алгоритм інтеграції прогнозної моделі у рекрутингові процеси організації

Інтеграція прогнозної моделі у рекрутингові процеси організації потребує чітко визначеної послідовності дій, яка дозволяє систематизувати процес прийняття рішень на основі прогнозів. Нижче представлено опис основних етапів взаємодії між учасниками процесу із зазначенням їхніх функцій.

1. Ініціювання процесу відкриття вакансії.

   Процес розпочинається з подання заявки на відкриття вакансії з боку керівника структурного підрозділу або менеджера проєкту. Цей крок передбачає формалізацію потреби у новому працівнику та передачу відповідної інформації до HR-відділу.
2. Первинна обробка заявки рекрутером.

   Після отримання запиту рекрутер або HR-менеджер вносить ключову інформацію до внутрішньої системи обліку вакансій (наприклад, HRIS), створюючи запис про нову позицію.
3. Формування параметрів для прогнозування

   На цьому етапі рекрутер або HR-аналітик заповнює обов’язкові атрибути вакансії, необхідні для роботи моделі. Серед них: категорія посади, очікуваний рівень досвіду кандидата, запропонований рівень оплати праці, передбачуване джерело пошуку кандидатів, а також сезон подання заявки.
4. Перевірка повноти даних.

Перед запуском моделі HR-аналітик перевіряє, чи всі необхідні змінні заповнено коректно. За відсутності повної інформації заявка повертається на попередній етап для доопрацювання.

5. Запуск прогнозної моделі.

   Після підтвердження коректності даних здійснюється обчислення прогнозу за допомогою моделі Random Forest. У результаті формується числове значення очікуваної тривалості закриття вакансії (у днях).

6. Інтерпретація прогнозу та погодження дій.

   Отриманий прогноз передається для опрацювання рекрутером і погоджується з керівником, який ініціював відкриття вакансії. На цьому етапі можуть обговорюватися зміни у графіку найму, розподілі обов'язків або терміновості заповнення позиції.

7. Прийняття рішення щодо рекрутингової стратегії.

   Враховуючи прогнозовану складність вакансії, рекрутер обирає відповідну стратегію пошуку: підключення додаткових каналів, залучення зовнішніх агентств, зміна вимог до кандидатів або адаптація умов вакансії.

8. Фіксація прогнозу для подальшого аналізу.

   Прогнозне значення зберігається у базі даних або HR-системі. Це дозволяє в майбутньому здійснювати порівняння з фактичним терміном закриття, що є основою для покращення точності моделі.

9. Перехід до стандартного процесу рекрутингу.

   Після погодження стратегії та фіксації прогнозу команда рекрутингу переходить до безпосереднього етапу підбору кандидатів.

4.3 Особливості реалізації та тестування моделі на практиці

Після побудови та оптимізації прогнозної моделі на основі алгоритму Random Forest було здійснено її апробацію в реальному бізнес-середовищі в межах науково-дослідної практики. Метою цього тестування стало перевірити, наскільки модель придатна до застосування у прикладних HR-сценаріях –

зокрема, для прогнозування термінів закриття нових вакансій одразу після їх відкриття. Для цього було використано нову вакансію, що на момент аналізу перебувала на етапі запуску в роботу:

- Категорія посади: Development
- Рівень досвіду: Middle
- Заробітна плата: 3000
- Кількість етапів відбору: 4
- Джерело пошуку кандидатів: LinkedIn
- Сезон відкриття: весна

У межах тестування модель була застосована для прогнозування терміну закриття цієї вакансії з метою демонстрації її потенційної прикладної цінності. Для застосування моделі було сформовано новий запис, що містить усі необхідні змінні, доступні на етапі відкриття вакансії. Дані введено вручну відповідно до формату навчального датасету, з урахуванням типів змінних.

```
> new_vacancy <- data.frame(
+     category = factor("Development", levels = levels(train_data$category)),
+     experience_level = factor("Middle", levels = levels(train_data$experience_level)),
+     salary = 3000,
+     hiring_steps = 4,
+     source = factor("LinkedIn", levels = levels(train_data$source)),
+     season = factor("spring", levels = levels(train_data$season))
+ )
```

Рисунок 22 – Додавання нової вакансії для тестування моделі.

До підготовленого запису було застосовано базову модель rf_model, що була навчена на історичних даних. У результаті отримано прогнозований термін закриття вакансії – 22 дні.

```
> predicted_days <- predict(rf_model, newdata = new_vacancy)
> cat("Прогнозований час закриття вакансії:", round(predicted_days), "днів\n")
Прогнозований час закриття вакансії: 22 днів
```

Рисунок 23 – Застосування створеної моделі.

Отриманий прогноз було представлено представникам компанії, яка виступала базою науково-дослідної практики, для ознайомлення з можливостями інструменту аналітичної підтримки рекрутингових рішень. Зокрема, було змодельовано сценарій, у якому прогнозне значення могло слугувати орієнтиром для погодження очікуваних строків найму між HR-відділом та керівником підрозділу. Таким чином, результат прогнозування розглядався як демонстраційний приклад практичного використання моделі в умовах реального бізнес-контексту.

З метою оцінки ефективності прогнозу заздалегідь було сформульовано орієнтовні критерії допустимого відхилення. Зокрема, якщо фактичний час закриття вакансії перебуватиме в межах ±5 днів від прогнозованого значення, результат вважатиметься прийнятним для практичного використання. У випадку значного розходження між очікуваним і реальним терміном закриття доцільно здійснити перегляд впливових факторів і потенційно внести корективи до структури моделі.

Після завершення рекрутингового процесу стало відомо, що фактичний термін закриття цієї вакансії склав 19 днів. Фактичне значення тривалості закриття вакансії (19 днів) відрізнялося від прогнозованого (22 дні) на 3 календарні дні, що знаходиться в межах прийнятного діапазону похибки для моделей такого типу. З огляду на варіативність зовнішніх і внутрішніх чинників, які впливають на швидкість найму, зазначене відхилення не може вважатися критичним. Навпаки, дещо завищене прогнозне значення є бажаним у практиці HR-планування, оскільки дозволяє уникнути ризиків, пов’язаних із недооцінкою тривалості процесу та можливими затримками реалізації кадрових потреб. Такий рівень точності підтверджує практичну придатність моделі для використання у реальному бізнес-середовищі, зокрема на етапі планування рекрутингових активностей.

4.4 Оцінка ефективності впровадження прогнозної моделі

Впровадження моделей прогнозної аналітики у сферу управління персоналом має на меті не лише удосконалення окремих етапів рекрутингового процесу, а й формування засад для прийняття більш обґрунтованих, стратегічно виважених рішень. У цьому контексті розроблена в межах дослідження модель прогнозування строків закриття вакансій розглядається як приклад прикладного інструменту, здатного забезпечити HR-фахівців додатковими аналітичними аргументами в умовах часової невизначеності. Її основне завдання – не замінити експертну оцінку, а підтримати її шляхом надання кількісних прогнозів, що ґрунтуються на аналізі історичних даних і виявлених закономірностях.

Актуальність подібного підходу зумовлена тим, що в реальній практиці процес планування строків закриття вакансій часто здійснюється на основі експертної оцінки, що може бути недостатньо обґрунтованою та нестабільною. За відсутності систематизованих інструментів прогнозування HR-відділи змушені орієнтуватися на інтуїцію або загальні середні значення, які не враховують специфіку конкретної позиції. Така ситуація може призводити до систематичного недооцінювання термінів найму, формування нереалістичних очікувань з боку керівництва та порушення проєктних дедлайнів у разі затримок у закритті ключових позицій.

Використання прогнозної моделі, побудованої на основі алгоритму Random Forest, дозволяє змінити підхід до планування – зробити його більш об’єктивним, адаптивним і прозорим. Модель забезпечує швидкий розрахунок орієнтовної тривалості найму на основі конкретних параметрів вакансії, що дає змогу оперативно адаптувати графіки, узгодити очікування між зацікавленими сторонами та зменшити імовірність кризових ситуацій, пов’язаних із браком персоналу. Відтак виникає потреба комплексного підходу до оцінки ефективності впровадження прогнозної моделі, який враховує не лише точність прогнозу, але й її вплив на якість управлінських рішень, передбачуваність процесів і загальну керованість HR-процесів.

Одним із ключових завдань у процесі впровадження прогнозної моделі в HR-практику є розробка системи критеріїв, що забезпечують об’єктивну оцінку її ефективності у прикладному контексті. Оскільки модель відіграє роль інструменту підтримки прийняття рішень, доцільно оцінювати не лише її прогностичну точність як таку, а й той організаційний ефект, який вона здатна забезпечити після інтеграції в реальні робочі процеси.

Серед першочергових критеріїв ефективності можна виокремити ступінь відповідності прогнозів фактичним результатам. При цьому оцінювання має здійснюватися не за одиничними прикладами, а системно – шляхом аналізу середнього абсолютного відхилення, частки прогнозів у межах допустимого діапазону похибки (наприклад, ±5 днів), а також стабільності моделі у розрізі різних категорій вакансій.

Другий важливий критерій – вплив моделі на якість планування. Зокрема, аналізу підлягає зниження частоти випадків, коли вакансії залишаються відкритими довше за запланований термін, або коли між HR-відділом та керівництвом виникають суперечності через невиправдані очікування. Успішне впровадження моделі має супроводжуватися зменшенням кількості таких помилок та підвищенням точності погоджених строків.

До критеріїв слід також віднести зміну якості комунікації між учасниками рекрутингового процесу. Наявність формалізованого прогнозу дозволяє зменшити суб’єктивізм у діалозі між рекрутерами та менеджерами, забезпечити єдину точку відліку при плануванні та посилити прозорість процесів.

Окремо можна виділити організаційні ефекти, які проявляються у покращенні ресурсного планування, підвищенні оперативності прийняття рішень, зростанні довіри до HR-аналітики та розширенні можливостей для побудови персоналізованих KPI для працівників, залучених до найму персоналу.

Для системної оцінки ефективності моделі у практиці доцільно застосовувати сукупність методів, що відображають як кількісні, так і якісні аспекти її функціонування:

- *Інструментальна оцінка* – аналіз точності прогнозів за допомогою формалізованих метрик (MAE, RMSE), які дозволяють кількісно оцінити результативність алгоритму.
- *Нормативна оцінка* – зіставлення фактичних значень із попередньо визначеними допустимими межами похибки (наприклад, ±5 днів), що забезпечує контроль відповідності результатів очікуванням.
- *Порівняльна оцінка* – порівняння результатів моделі з альтернативними підходами до планування або іншими аналітичними алгоритмами, що дає змогу оцінити її відносну ефективність.
- *Експертна оцінка* – якісне оцінювання моделі з боку її користувачів (HR-аналітиків, менеджерів), яке дозволяє виявити практичні переваги чи обмеження при реальному застосуванні.
- *Суб'єктивна або поведінкова оцінка* – аналіз змін у поведінці користувачів, рівня довіри до моделі та готовності приймати прогнози як частину регулярного планування.

Поєднання цих підходів дає змогу сформувати багатовимірну оцінку впливу моделі не лише як аналітичного інструменту, а й як елемента трансформації управлінських практик у сфері підбору персоналу.

4.5 Стратегії розширення аналітичного застосування моделі в бізнес-середовищі

З огляду на результати тестового застосування та потенціал моделі у підвищенні точності та передбачуваності кадрового планування, доцільно розглянути можливості її подальшого масштабування в межах HR-функції. Побудована модель прогнозування строків закриття вакансій є лише першим кроком до створення більш комплексної аналітичної екосистеми, яка підтримуватиме прийняття рішень на основі даних у ширшому спектрі задач управління персоналом.

Одним із напрямів подальшого розвитку є використання моделі як інструменту для оптимізації операційного планування в межах HR-відділу. Зокрема, модель може слугувати базою для оцінювання індивідуального навантаження на кожного рекрутера з урахуванням прогнозованої тривалості закриття вакансій. Такий підхід дозволяє розподіляти потік відкритих позицій більш збалансовано, враховуючи не лише кількість заявок, але й складність та часову тривалість кожного процесу найму. Це створює передумови для персоналізованого управління ефективністю, коли метрики результативності (KPI) формуються не лише на основі кількісних показників, але й з урахуванням прогнозного обсягу роботи. Наприклад, рекрутер, який закриває кілька вакансій із високою передбачуваною тривалістю, не має оцінюватися за тією ж шкалою, що й спеціаліст, відповідальний за оперативні найми.

Крім того, аналітична модель може бути використана як інструмент підтримки стратегічного планування. Знання про те, які категорії вакансій найімовірніше вимагатимуть більше часу для закриття, дає змогу HR-департаменту заздалегідь погоджувати кадрові потреби з іншими підрозділами, розробляти альтернативні сценарії залучення персоналу, а також проактивно планувати внутрішню мобільність або перекваліфікацію. Таким чином, прогнозна аналітика стає не лише частиною виконання операційних задач, але й чинником підвищення адаптивності організації до змін у кадровому середовищі.

Важливим елементом масштабування є інтеграція моделі в наявну цифрову інфраструктуру компанії. Повноцінне впровадження у середовище HRIS або систему підтримки прийняття рішень (DSS) дозволить автоматизувати процес формування прогнозів. Зокрема, за умови наявності структурованих даних про вакансії в електронному вигляді, можливо реалізувати механізм автоматичного виклику моделі одразу після створення нової позиції. Отриманий прогноз може бути збережений у профілі вакансії, доступний для аналітичних звітів, або надісланий відповідальному рекрутеру як орієнтовний дедлайн.

Інтеграція моделі у HRIS відкриває можливість регулярного оновлення алгоритму на основі накопичення нових спостережень, без потреби в ручному

втручанні. Це відповідає принципу побудови адаптивних систем, які підлаштовуються до змін у внутрішніх процесах та зовнішньому середовищі. Така здатність до самовдосконалення є однією з ключових вимог до сучасних аналітичних рішень у сфері управління персоналом.

Окрім автоматизації, впровадження моделі може сприяти підвищенню аналітичної зрілості організації. За наявності інструменту, який формує кількісні прогнози на основі історичних закономірностей, HR-команда переходить до практики прийняття рішень, базованих на даних. Це, у свою чергу, підвищує довіру з боку керівництва до аналітики, зменшує суб'єктивізм у комунікації, а також покращує узгодженість між планами рекрутингу та операційною діяльністю компанії.

З урахуванням наведеного, варто зазначити, що подальший розвиток моделі повинен супроводжуватись як технологічними, так і організаційними змінами. З технічної точки зору – це вдосконалення алгоритму, врахування нових змінних, підвищення продуктивності. З боку управління – це впровадження політик, які забезпечують регулярне оновлення даних, розподіл відповідальності за аналітичні результати, а також навчання персоналу роботі з такими інструментами. У сукупності це створює умови для побудови сталої аналітичної практики в HR-напрямі, здатної реагувати на виклики ринку та підтримувати стратегічні цілі організації.

Висновки

У четвертому розділі було розглянуто практичні аспекти впровадження розробленої моделі прогнозування термінів закриття вакансій у реальні бізнес-процеси. Проведений аналіз дозволяє зробити низку висновків щодо потенціалу моделі як ефективного інструменту підтримки прийняття управлінських рішень у сфері управління персоналом.

По-перше, сформульовано концепцію використання моделі як аналітичного ядра системи підтримки прийняття рішень (DSS), що інтегрується

у функціонування HR-відділу без потреби у складних технічних рішеннях. Завдяки гнучкості реалізації, модель може бути вбудована у вже існуючі системи (наприклад, HRIS або Excel-інтерфейси), що забезпечує її доступність для організацій різного масштабу. Ключовими користувачами прогнозної системи виступають HR-аналітики, рекрутери, менеджери підрозділів та керівництво компанії, кожен з яких отримує специфічну аналітичну вигоду від використання прогнозу.

По-друге, представлено алгоритм інтеграції моделі у рекрутингові процеси, що охоплює повний цикл: від ініціювання вакансії до вибору стратегії пошуку на основі прогнозного значення. Така структуризація взаємодії між учасниками процесу дозволяє перетворити прогнозування на системний елемент рекрутингової діяльності, а не на епізодичну практику.

По-третє, результати практичного тестування моделі свідчать про її достатній рівень точності у прикладних умовах. Прогноз, сформований для реальної вакансії, відхилився від фактичного терміну закриття лише на три дні, що відповідає встановленим критеріям прийнятної похибки. Такий рівень точності є достатнім для використання моделі в задачах тактичного планування.

По-четверте, модель має потенціал масштабування в межах організації. Вона може застосовуватись для оцінки навантаження на рекрутерів, формування реалістичних планів найму та стратегічного управління персоналом. Інтеграція в цифрову інфраструктуру забезпечує автоматизацію прогнозування, оновлення алгоритму та адаптивність системи.

Нарешті, впровадження прогнозної моделі сприяє зростанню аналітичної зрілості HR-функції та трансформації управлінських практик на основі даних. За умови дотримання принципів регулярного оновлення, навчання персоналу та систематичного моніторингу ефективності, модель може стати основою для формування сталої аналітичної екосистеми в управлінні персоналом. Таким чином, розроблене рішення має не лише прикладну цінність, а й значний стратегічний потенціал у контексті цифрової трансформації HR-процесів.

ЗАГАЛЬНІ ВИСНОВКИ

У межах дипломної роботи було з’ясовано можливості застосування методів Data Science для вирішення актуального завдання у сфері управління персоналом – прогнозування термінів закриття вакансій. Актуальність теми дослідження обумовлена зростаючою потребою в автоматизації HR-процесів та підвищенні ефективності рекрутингових стратегій шляхом впровадження аналітичних інструментів, що ґрунтуються на сучасних алгоритмах машинного навчання. У роботі було визначено, що традиційні підходи до оцінки ефективності рекрутингу не враховують часові параметри процесу найму, хоча саме ці характеристики мають вирішальне значення для забезпечення безперервності бізнес-процесів, планування ресурсів та мінімізації фінансових втрат, пов’язаних із вакантними позиціями.

У результаті огляду наукових джерел виявлено, що попри зростання інтересу до HR-аналітики, питання прогнозування тривалості рекрутингового циклу все ще залишаються недостатньо опрацьованими. У зв’язку з цим було запропоновано використати алгоритм Random Forest для побудови моделі, здатної передбачати кількість днів, необхідних для закриття вакансії, на основі рекрутингових параметрів. Вибір цього методу зумовлений його здатністю працювати з гетерогенними наборами змінних, високою точністю прогнозу та інтерпретованістю результатів, що є важливою умовою в управлінських практиках.

У процесі дослідження здійснено всі етапи аналітичної роботи з даними, включаючи їх підготовку, вибір релевантних змінних, моделювання та оцінювання точності отриманих результатів. До складу змінних було включено такі характеристики, як категорія вакансії, рівень досвіду, канал залучення, сезонність та інші чинники, які потенційно можуть впливати на тривалість найму. Дані були попередньо очищені, нормалізовані та оброблені з урахуванням типу змінних, після чого застосовано просте випадкове розділення даних на навчальну та тестову підвибірки за допомогою функції sample() для тренування

та тестування моделі. Алгоритм Random Forest показав задовільну точність прогнозу, продемонструвавши середню абсолютну похибку (MAE) на рівні близько 5 днів, що є прийнятним показником для задач кадрового планування в реальних умовах.

Крім побудови прогнозної моделі, у роботі було здійснено аналіз важливості змінних, що дозволило ідентифікувати ключові фактори впливу на терміни закриття вакансій. Зокрема, виявлено, що найбільший вплив мають категорія посади, рівень кваліфікації кандидата, а також сезон, у якому була відкрита вакансія. Також було зафіксовано значущу роль каналу залучення персоналу – вакансії, опубліковані через внутрішні ресурси компанії, як правило, закриваються швидше, ніж ті, що поширюються через зовнішні платформи. Ці висновки можуть бути використані для вдосконалення стратегій підбору персоналу, визначення пріоритетів у розміщенні вакансій та оптимізації часу на найм.

Результати дослідження підтверджують, що інтеграція моделей прогнозування у практику HRM сприяє більш раціональному використанню ресурсів компанії, дозволяє уникати затримок у формуванні команд та створює передумови для підвищення загальної конкурентоспроможності організації. Крім того, обґрунтовано можливість масштабування побудованої моделі на інші бізнес-процеси, зокрема планування навантаження на рекрутерів, оцінку ризиків незакриття критичних позицій у визначені строки.

Практична цінність розробленого підходу полягає у створенні інструменту, який може бути інтегрований у внутрішні HR-аналітичні системи компаній, працювати на основі наявних даних і забезпечувати підтримку управлінських рішень у реальному часі. Використання мови програмування R та спеціалізованих бібліотек (зокрема randomForest, caret, ggplot2) дозволило реалізувати модель у вигляді програмного прототипу, придатного до подальшої адаптації під конкретні потреби організації. У процесі реалізації було створено скрипти для обробки даних, тренування моделі, оцінки її якості та побудови візуалізацій, що забезпечує повну відтворюваність результатів дослідження.

За результатами тестування моделі встановлено, що середня абсолютна похибка прогнозу (MAE) склала 5.13 днів, а корінь середньої квадратичної похибки (RMSE) – 7.18 днів, що відповідає прийнятному рівню точності для завдань HR-планування. У типових прикладних кейсах відхилення між прогнозованими та фактичними строками закриття вакансій не перевищувало трьох днів, а частка прогнозів у межах допустимого інтервалу ±5 днів сягала орієнтовно 70%. На основі експертного аналізу результатів було зафіксовано, що застосування моделі дозволило підвищити точність планування строків закриття вакансій приблизно на 35% порівняно з традиційною практикою. Підвищення точності оцінювалося як зменшення середнього відхилення між плановими та фактичними строками у порівнянні з експертним прогнозуванням, виражене у відсотках, що підтверджує практичну доцільність застосування моделі для підвищення передбачуваності HR-процесів.

Таким чином, виконана дипломна робота підтвердила можливість ефективного застосування методів Data Science у сфері управління персоналом для вирішення задач прогнозного характеру. Поставлена мета дослідження – розробка моделі прогнозування термінів закриття вакансій – була досягнута повністю. Усі завдання, що випливали з дослідницької логіки, реалізовано: проведено аналіз літератури, визначено методологію, побудовано і протестовано модель, проаналізовано результати, зроблено висновки щодо подальшого практичного застосування. Отримані результати відкривають перспективи для подальших досліджень у сфері прогнозної HR-аналітики, зокрема щодо адаптації моделей до різних типів компаній, розширення переліку змінних, а також використання більш складних алгоритмів у задачах високої варіативності рекрутингових процесів.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bahuguna P., Srivastava R., Tiwari S. Human resources analytics: where do we go from here. *Benchmarking: An International Journal*. 2023. Vol. 31, № 2. P. 640-668.
2. Balcioglu Y., Artar M., Erdil O. Data-Driven Insights into HR Recruitment Performance: A Machine Learning Approach Based on Real-World Data from an Italian Security Company. *Journal Of Current Debates In Social Sciences (CUBES).* 2023. P. 282-293. URL: https://www.researchgate.net/publication/375027397_Data-Driven_Insights_into_HR_Recruitment_Performance_A_Machine_Learning_Approach_Based_on_Real-World_Data_from_an_Italian_Security_Company (Last accessed: 28.02.2025).
3. Belizón M. J., Majarín D., Aguado, D. Human resources analytics in practice: A knowledge discovery process. *European Management Review*. 2023. Vol. 21, № 1. P. 1–19. DOI: 10.1111/emre.12605.
4. Biau, G., Scornet, E. A Random Forest Guided Tour. *TEST*. 2016. 25(2), 197–227. DOI: 10.1007/s11749-016-0481-7.
5. California Consumer Privacy Act. 2018. URL: https://oag.ca.gov/privacy/ccpa (Last accessed: 10.01.2025).
6. Data Protection Act 2018. URL: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted (Last accessed: 10.01.2025).
7. Delen D., Zolbanin H. The analytics paradigm in business research. *Journal of Business Research*. 2018. Vol. 90. P. 186-195.
8. Fernando J., Baldelovar M. *Decision Support System: Overview, Different Types and Elements*. Technoarete Transactions on Intelligent Data Mining and Knowledge Discovery. 2022. Vol.2. DOI: 10.36647/TTIDMKD/02.02.A003.
9. Frazzetto P., Haq M.U., Sperduti A. Enhancing Human Resources through Data Science: a Case in Recruiting. *ITADATA2023: The 2nd Italian Conference on Big Data and Data Science Proceedings.* 2023. URL: https://ceur-ws.org/Vol-3606/paper71.pdf (Last accessed: 20.02.2025).

10. Glennie M., Buick F., Blackman D., Weeratunga V., Bertuol M., West D., Dickinson H. Opportunities and challenges of using workforce big data: Insights from a mixed methods study on flexible working. *Australian Journal of Public Administration*. 2023. Vol. 82. P. 590 – 595. DOI: 10.1111/1467-8500.12591.
11. Goodfellow I., Bengio Y., Courville A. "Deep Learning". Cambridge: MIT Press, 2016. 775 p.
12. Guenole N., Ferrar J., Feinzig S., The power of people: Learn how successful organizations use workforce analytics to improve business performance. FT Press. 2017. 315 p.
13. Gupta S., Sharma R. Types of HR Analytics Used for the Prediction of Employee Turnover in Different Strategic Firms with the use of Enterprise Social Media. *Proceedings of the 5th European International Conference on Industrial Engineering and Operations Management*. 2023. P. 1977–1994. URL: https://www.researchgate.net/publication/368313154_Types_of_HR_Analytics_Used_for_the_Prediction_of_Employee_Turnover_in_Different_Strategic_Firms_with_the_use_of_Enterprise_Social_Media.
14. Guseva Yu., Kazarova O., Dumanska I., Gorodetskyy M., Melnichuk L., Saienko V. Data Protection Policy Impact on the Company Development. W*SEAS Transactions On Environment And development.* 2022. Vol. 18. P. 232-246. DOI: 10.37394/232015.2022.18.25.
15. Hamieddine C., Tigani S., Akioud M., Saadane R., Chehri A. From Data to Decisions: Exploring Data Analytics in HR for Agile Organizational Decision Making. *Procedia Computer Science.* 2024. Vol. 246. P. 4901-4908. DOI: 10.1016/j.procs.2024.09.446.
16. Harding J. Data Quality in The Integration and Analysis of Data From Multiple Sources: Some Research Challenges. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 2013. Vol. 2, № 1. P. 59-63. URL: https://www.researchgate.net/publication/289558449_Data_quality_in_the_integration_and_analysis_of_data_from_multiple_sources_Some_research_challenges (Last accessed: 09.01.2025).

17. Hassan M. B., Prediction of increased attrition rate of employee using people analytics: Master's thesis. Lappeenranta-Lahti University of Technology. 2022. URL: https://lutpub.lut.fi/bitstream/handle/10024/164000/mastersthesis_hassasn_muhammadbilal.pdf (Last accessed: 12.02.2025).
18. James G., Witten D., Hastie T. and Tibshirani R. An Introduction to Statistical Learning: With Applications in R. 2nd Edition. 2021. Springer, Berlin. https://doi.org/10.1007/978-1-0716-1418-1.
19. Jayanti L D., Wasesa M. Application of Predictive Analytics To Improve The Hiring Process In A Telecommunications Company. *Jurnal CoreIT*. 2022. Vol. 8, № 1. P. 32-39. DOI: 10.24014/coreit.v8i1.16915.
20. Jiang E., Zheng J., Zhu J. The Statistical Analysis of HR Employee Retention, Salary Variation of Remote Work and Earthquake Occurrence Probability. *BCP Business & Management*. 2023. Vol. 38. P. 2731-2737. DOI: 10.54691/bcpbm.v38i.4182.
21. Jogarao M., Hemalatha T., & Naidu, S. T. Leveraging HR Analytics for Data-Driven Decision Making: A Comprehensive Review. *IJFANS International Journal of Food and Nutritional Sciences.* 2022. Vol 11, № 10. 1774-1784. DOI: 10.13140/RG.2.2.16977.30562.
22. Keary B., Perry K. Possibility Theory Quantification in Human Capital Management: A Scientific Machine Learning (SciML) Perspective. *arXiv (Cornell University).* DOI: 10.48550/arXiv.2302.14088.
23. Lepenioti K., Bousdekis A., Apostolou D., Mentzas G. Prescriptive analytics: literature review and research challenges. *International Journal of Information Management*. 2020. Vol. 50. P. 57–70.
24. Müller O., Junglas I., Debortoli, S., Brocke J. Using Text Analytics to Derive Customer Service Management Benefits from Unstructured Data. *MIS Quarterly Executive*. 2016. Vol. 15, № 4. P. 243-258.
25. Natekin, A., & Knoll, A. Gradient Boosting Machines, A Tutorial. *Frontiers in Neurorobotics*. 2013. Vol.7, Article 21. https://doi.org/10.3389/fnbot.2013.00021.
26. Panda D., Srikanth Ch., Bharadwaj K., Chandra K. Employee Monitoring System Using Power BI. *International Journal for Research in Applied Science and*

Engineering Technology. 2024. Vol. 12. P. 1942-1948. DOI: 10.22214/ijraset.2024.60233.

27. Pandey D., Kumar C., Singh A., Umbarkar D., Sharma S. AI-Powered Recruitment: Transforming Talent Acquisition in the Digital Age. *Journal of Informatics Education and Research*. 2025. Vol. 5, № 1. P. 1742-1750.
28. Pessach D., Singer G., Avrahami D., Ben-Gal H.C., Shmueli E., Ben-Gal I. Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *Decision Support Systems*. 2020. Vol. 134. DOI: 10.1016/j.dss.2020.113290.
29. Quan T. Z., Raheem M. Human Resource Analytics on Data Science Employment Based on Specialized Skill Sets with Salary Prediction. *International Journal of Data Science*. 2023. Vol. 4, № 1. P. 40-59.
30. Ranjan N., Prasad R. Text Analytics: An Application of Text Mining. *Journal of Data Mining and Management*. 2022. Vol. 6, № 3. P. 1-6. DOI: 10.46610/JoDMM.2021.v06i03.001.
31. Rangineni S., Bhanushali A., Suryadevara M., Venkata S., Peddireddy K. A Review on Enhancing Data Quality for Optimal Data Analytics Performance. *International Journal of Computer Sciences and Engineering*. 2023. Vol. 11, № 10. P. 51-58. DOI: 10.26438/ijcse/v11i10.5158.
32. Technical and organizational measures for the protection of personal data : Data Protection Act. 2018. URL: https://www.paragon.world/system/files/2023-03/de_cc_TOMs_PCC_Czechia_ENG_2020-07-20.pdf
(Last accessed: 11.01.2025).
33. Tuli F.A., Sachani D.K., Venapusa S.R. The Role of HR Analytics in Strategic Decision Making: Leveraging Data for Talent Management. *Asian Business Review*. 2024. Vol. 14, № 1. P. 31-42. DOI: 10.52783/jier.v4i2.1143.
34. Wahyuni H. Big Data Analysis in Human Resources Decision Making: Optimizing Workforce Management. *Jurnal Riset Manajemen Sains Indonesia*. 2024. Vol.15, № 1. P. 58-69. DOI: 10.21009/JRMSI.015.1.06.

35. Zavgorodniy A. How To Use Data Science in The HR Industry. *Medium*: website. URL: https://medium.com/unicsoft/how-to-use-data-science-in-the-hr-industry-faeb30c14460 (Last accessed: 19.01.2025).
36. Мінакова В. П., Шіковець К. О. Актуальність використання моделі Big Data в бізнес-процесах. *Економіка і Суспільство*. 2017. Вип. 10. С. 892–896.
37. Про захист персональних даних: Закон України від 01 черв. 2010 р. № 2297-VI. URL: https://zakon.rada.gov.ua/laws/show/2297-17#Text (дата звернення: 21.01.2025).