

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерних наук _____
(повна назва)

Кафедра _____ програмної інженерії _____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти _____ другий (магістерський) _____

_____ Дослідження методів аналізу дистанційної роботи працівників _____
_____ Рекомендації для підвищення ефективності _____
(тема)

Виконала:

Здобувачка _____ 2 _____ року навчання
групи _____ ПЗМ-23-4 _____

_____ Юлія ПАРХОМЕНКО _____

(власне ім'я, прізвище)

Спеціальність _____ 121 – Інженерія програмного
забезпечення _____

(код і повна назва спеціальності)

Тип програми _____ освітньо-наукова _____

Керівник _____ доц. Володимир ЛЕЩИНСЬКИЙ _____

(посада, власне ім'я, прізвище)

Допускається до захисту
Зав. кафедри _____

(підпис)

_____ Кирило СМЕЛЯКОВ _____

(Власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерних наук
 Кафедра _____ програмної інженерії
 Рівень вищої освіти _____ другий (магістерський)
 Спеціальність _____ 121 – Інженерія програмного забезпечення
 Тип програми _____ освітньо-наукова програма
 Освітня програма _____ Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

«___» _____ 2025 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентці _____ Пархоменко Юлії Юріївні
 (прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження методів аналізу дистанційної роботи працівників. Рекомендації для підвищення ефективності»

Затверджена наказом по університету від 15.04.2025 р. №290 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 16.06.2025

3. Вихідні дані до роботи:

У дослідницькій роботі передбачити: дані з відкритого набору Kaggle про ефективність працівників у різних форматах зайнятості, набір IBM HR Analytics для оцінки плинності кадрів; інструменти машинного навчання (Random Forest, KMeans) та бібліотеки Python (Pandas, Scikit-learn, Flask) для обробки даних і побудови моделей; програмну реалізацію веб-застосунку; пояснювальну записку

4. Перелік питань, що потрібно опрацювати в роботі:

Аналіз предметної галузі, огляд та аналіз наукових джерел, постановка задачі, теоретичне дослідження, висновки, перелік посилань, додатки

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	16.04.2025	<i>виконано</i>
2	Аналіз предметної галузі і постановка задачі	17.04.2025 – 18.04.2025	<i>виконано</i>
3	Огляд й аналіз літературних, наукових джерел	19.04.2025 – 23.04.2025	<i>виконано</i>
4	Постановка задачі	24.04.2025 – 26.04.2025	<i>виконано</i>
5	Теоретичне дослідження	27.04.2025 – 06.05.2025	<i>виконано</i>
6	Підготовка до апробації результатів дослідження. Публікація матеріалів	07.05.2025 – 13.05.2025	<i>виконано</i>
7	Аналіз отриманих результатів та формування висновків	14.05.2025 – 15.05.2025	<i>виконано</i>
8	Підготовка пояснювальної записки	16.05.2025 – 29.05.2025	<i>виконано</i>
9	Підготовка презентації та доповіді	30.05.2025 – 03.06.2025	<i>виконано</i>
10	Перевірка на плагіат	04.06.2025	<i>виконано</i>
11	Нормоконтроль	05.06.2025	<i>виконано</i>
12	Рецензування	13.06.2025	<i>виконано</i>
13	Попередній захист	14.06.2025	<i>виконано</i>
14	Занесення диплома в електронний архів	16.06.2025	<i>виконано</i>
15	Допуск до захисту у зав. кафедри	17.06.2025	<i>виконано</i>

Дата видачі завдання 16 квітня 2025р.

Здобувачка



(підпис)

Юлія ПАРХОМЕНКО

Керівник роботи

доц. Володимир ЛЕЩИНСЬКИЙ

(підпис)

(посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить: 64 с., 12 рис., 2 формули, 5 табл., 14 джерел.

АНАЛІТИЧНА СИСТЕМА, ДАНІ HR-АНАЛІТИКИ, ДИСТАНЦІЙНА РОБОТА, КЛАСТЕРИЗАЦІЯ, МОДЕЛІ МАШИННОГО НАВЧАННЯ, ПІДТРИМКА ПРИЙНЯТТЯ РІШЕНЬ, ПРОДУКТИВНІСТЬ, РИЗИК ЗВІЛЬНЕННЯ, ТРУДОВА ЕФЕКТИВНІСТЬ.

У роботі досліджено проблему оцінювання продуктивності працівників і прогнозування ризику їх звільнення в умовах дистанційної зайнятості. Актуальність дослідження зумовлена широким переходом компаній до гнучких моделей роботи, що створює запит на ефективні інструменти підтримки прийняття управлінських рішень на основі HR-даних.

Об'єктом дослідження є процеси аналізу трудової ефективності персоналу в умовах дистанційної праці. Предметом дослідження виступають аналітичні моделі та методи машинного навчання для виявлення закономірностей і прогнозування результативності працівників.

Метою роботи є проведення дослідження впливу факторів, що визначають ефективність дистанційної праці, та розробка аналітичної моделі для оцінки продуктивності й ризику звільнення на основі HR-даних.

У роботі використано методи кластерного аналізу, нормалізації даних, регресійного та класифікаційного моделювання. Проведено генерацію додаткових оцінкових ознак, побудовано кілька моделей із використанням алгоритмів Random Forest, KMeans та GridSearchCV. Застосовано Python, Flask, Pandas, Scikit-learn.

Результатом дослідження є створена аналітична система з вебінтерфейсом, що дозволяє проводити індивідуальний аналіз працівників і формувати управлінські рекомендації. У межах дослідження розроблено також модуль пояснення прогнозів, що підвищує інтерпретованість результатів.

Розроблена система може бути інтегрована у внутрішні HR-платформи організацій або використана як самостійний аналітичний інструмент. Вона

орієнтована на управлінців та HR-фахівців, що прагнуть приймати рішення на основі даних.

Результати роботи можуть бути використані в IT-компаніях, освітніх установах, а також у сфері управління персоналом для формування стратегій утримання кадрів. Дослідження має соціально-економічне значення, оскільки сприяє підвищенню ефективності організацій та добробуту працівників.

ANALYTICAL SYSTEM, HR ANALYTICS DATA, REMOTE WORK, CLUSTERING, MACHINE LEARNING MODELS, DECISION SUPPORT, PRODUCTIVITY, TURNOVER RISK, WORKFORCE EFFICIENCY.

This research explores the issue of assessing employee productivity and predicting attrition risk under remote work conditions. The relevance of the study is driven by the increasing prevalence of flexible work formats, which necessitate the development of effective decision support tools based on HR analytics data.

The object of the study is the process of analyzing employee performance in the context of remote work. The subject of the research includes analytical methods and machine learning models for predicting employee productivity and turnover risk.

The aim of the study is to conduct an in-depth investigation into the effectiveness of remote work using HR analytics data and to develop a model for evaluating employee productivity and attrition risk.

The study uses statistical methods and machine learning techniques, including data normalization, clustering, regression, and classification. To compensate for missing indicators in the dataset, synthetic data were generated based on logical dependencies between available variables. The system was implemented using Python and libraries such as Pandas, Scikit-learn, and Flask, and includes a user-friendly web interface.

The result of the work is a developed analytical system that enables individual employee analysis, visual performance interpretation, and automatic generation of recommendations for HR decisions. A module for interpreting model predictions has also been developed, which enhances transparency and trust in analytical results.

The developed solution can be integrated into corporate HR platforms and used as a decision support tool by managers and HR professionals. The research has a strong practical orientation and can be applied in various organizations seeking to increase operational efficiency and employee well-being.

Завідувачу кафедри
ПІ
(скорочена назва кафедри)
проф. Кирилу СМЕЛЯКОВУ
(вчене звання, сласне ім'я, прізвище)

ЗАЯВА

щодо самостійності виконання кваліфікаційної роботи та можливості її публікації
(та/або публікації анотації кваліфікаційної роботи) в електронному архіві
відкритого доступу EIAr KhNURE

Я, Пархоменко Юлія Юріївна, студентка гр. ПЗм-23-4, здобувачка вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів аналізу дистанційної роботи працівників. Рекомендації для підвищення ефективності», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомена з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

25.05.2025

Юлія ПАРХОМЕНКО

ЗМІСТ

Вступ	10
1 Аналіз предметної галузі	12
1.1 Сучасний стан дистанційної роботи в ІТ-галузі	12
1.2 Постановка задачі	14
2 Огляд й аналіз літературних та наукових джерел	17
2.1 Аналіз глобальних наукових джерел	17
2.2 Аналіз наукових джерел університету	18
2.3 Узагальнення підходів до прогнозування продуктивності та ризику звільнення	19
3 Проектні рішення і архітектура системи	21
3.1 Загальна архітектура програмної системи	21
3.2 Структура веб-застосунку	22
3.3 Взаємодія компонентів та схема використання моделей	24
4 Опис програмної реалізації та методів машинного навчання	27
4.1 Підготовка навчальних даних та обробка ознак	27
4.2 Вибір моделей та обґрунтування алгоритмів	29
4.3 Інтерпретація результатів та формування рекомендацій	30
5 Експериментальні дослідження	33
5.1 Характеристика вхідних даних та підготовка до тестування	33
5.2 Аналіз результатів та приклади HR-сценаріїв	35
Висновки	43
Перелік джерел посилання	45
Перелік джерел посилання за науковими напрямками керівника та науковців кафедри програмної інженерії	47
Додаток А Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ	48
Додаток Б Слайди презентації	50
Додаток В Апробація результатів роботи	57

Додаток Д Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015 64

ВСТУП

Сучасний ринок праці стрімко змінюється під впливом цифрової трансформації, глобалізації та поширення гнучких форм зайнятості. Одним із найпомітніших трендів останніх років є широке впровадження дистанційної роботи, що набуло особливої актуальності у зв'язку з пандемією COVID-19. Ця трансформація створила нові виклики для роботодавців у питаннях оцінки ефективності праці, підтримки мотивації та зниження плинності кадрів.

Водночас, сучасні інформаційні технології відкривають нові можливості для збору, аналізу та інтерпретації даних про працівників. HR-аналітика, як напрям, що базується на системному використанні даних для прийняття управлінських рішень, дає змогу побудувати об'єктивну оцінку ефективності персоналу та передбачати потенційні ризики звільнення. Особливої значущості набуває застосування методів машинного навчання та кластерного аналізу, які дозволяють виявляти приховані закономірності у поведінці працівників.

Актуальність теми обумовлена необхідністю розробки інструментів, які допоможуть компаніям адаптуватися до нових умов роботи, приймати обґрунтовані рішення щодо управління персоналом і підтримувати високий рівень продуктивності. Науковий інтерес полягає у дослідженні можливостей інтелектуального аналізу HR-даних для виявлення залежностей між добробутом працівника, його робочим навантаженням та ефективністю.

Об'єктом дослідження є процеси прийняття управлінських рішень на основі HR-аналітики в умовах дистанційної зайнятості.

Предметом дослідження виступають методи обробки та аналізу даних про працівників для оцінки продуктивності та прогнозування ризику звільнення.

Метою роботи є проведення дослідження трудової ефективності працівників за умов дистанційної роботи та розробка аналітичної системи підтримки прийняття рішень на основі HR-даних.

Для досягнення поставленої мети в роботі було вирішено такі завдання:

- здійснити аналіз сучасного стану досліджень у сфері HR-аналітики та прогнозування ефективності працівників;

- підготувати й попередньо обробити дані про дистанційних працівників;
- сформуванати набір ознак для оцінювання добробуту, задоволеності та навантаження;
- побудувати модель кластеризації типів працівників;
- створити прогностичні моделі продуктивності та ризику звільнення;
- реалізувати інтерактивну веб-систему аналітики із поясненнями результатів;
- оцінити ефективність запропонованих моделей і сформулювати рекомендації щодо впровадження.

Практична цінність роботи полягає у створенні комплексного інструменту, який може бути використаний у корпоративному середовищі для моніторингу стану працівників, попередження вигорання та зниження плинності кадрів.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ І ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз предметної галузі дослідження

Дистанційна робота в ІТ-секторі за останні п'ять років зазнала революційних змін. Якщо у 2019 році лише 28% ІТ-компаній пропонували віддалену роботу як опцію, то після пандемії COVID-19 цей показник зріс до 68% [1]. Проте у 2023-2024 роках спостерігається певна корекція: частка повністю віддалених працівників скоротилася з 34% до 29%, тоді як гібридний формат зріс з 31% до 36%. При цьому 68% працівників оцінюють свій досвід віддаленої роботи як «дуже позитивний» (див. рис. 1.1).

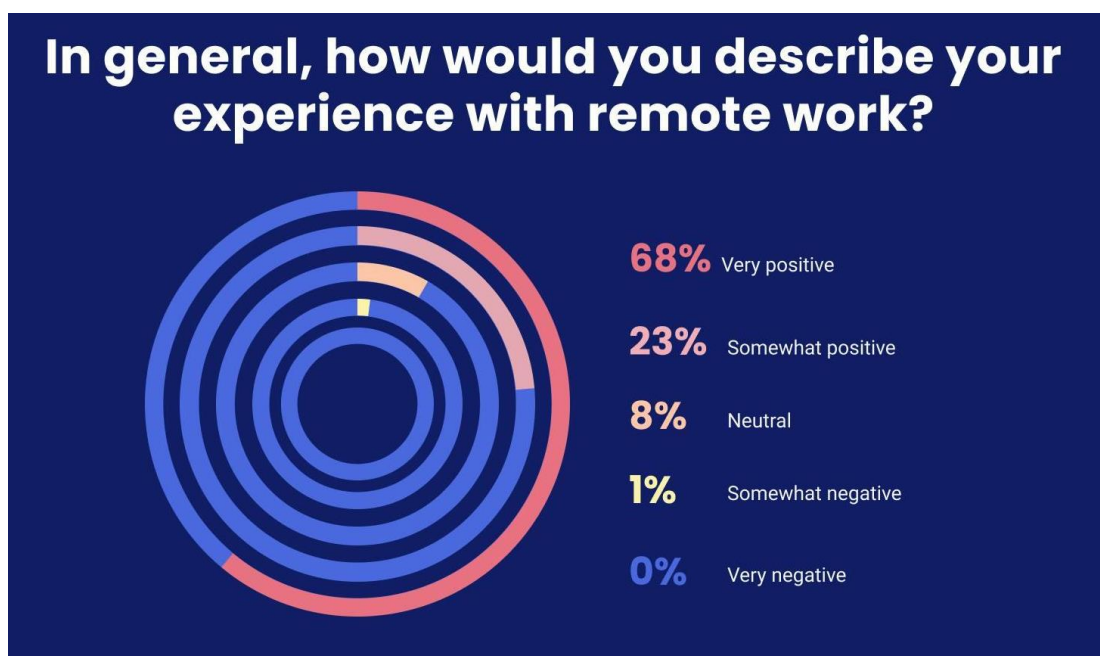


Рисунок 1.1 – Досвід працівників у дистанційному форматі роботи (за даними [1])

Фактори, що впливають на динаміку:

- технологічний прогрес;
- зміна очікувань працівників;
- економічні фактори.

Ця трансформація стала можливою завдяки стрімкому розвитку хмарних технологій, інструментів спільної роботи (наприклад, Jira, GitHub) та автоматизації процесів.

Технологічний прогрес, а саме розвиток хмарних технологій (AWS, Google Cloud) та інструментів колаборації (Figma, Miro), зробив віддалену роботу технічно доцільною. Проте масштабний перехід виявив і суттєві виклики: від кібербезпеки до психологічного вигорання.

Зміна очікувань працівників стала ключовим чинником трансформації ринку праці. Згідно з опитуванням Buffer «State of Remote Work», 82% IT-фахівців вважають можливість працювати віддалено критично важливою при виборі роботодавця, а 67% готові погодитися на зниження зарплати на 5-10% для збереження гібридного формату. Ці дані підтверджуються дослідженням Owl Labs, де 76% працівників IT-сфери зазначили, що віддалена робота покращила їхній work-life баланс [2].

Економічні фактори також відіграли вирішальну роль. Згідно з аналізом FlexJobs, IT-компанії економлять до \$11,000 на одного віддаленого працівника щорічно за рахунок:

- скорочення витрат на офісні приміщення (30-50%);
- зниження витрат на обладнання (15-25%);
- відмова від релокаційних пакетів (20-35%) [3].

Проте масштабний перехід до дистанційних форматів виявив нові виклики. Звіт IBM Security зазначає, що 45% кібератак націлені на віддалених працівників через слабкі точки в домашніх мережах [4]. Крім того, дослідження Stanford University (2023) виявило, що 58% IT-спеціалістів відчувають ізоляцію, а 41% стикаються з емоційним вигоранням через надлишок онлайн-зустрічей.

Для подолання цих проблем провідні компанії, такі як Microsoft та Salesforce, інвестують у:

- віртуальні офіси (платформи Gather, SpatialChat);
- програми психологічної підтримки (доступ до терапевтів через додатки Calm, Headspace);
- автоматизовані інструменти аналітики (наприклад, Microsoft Viva Insights) для моніторингу навантаження.

Прогнози вказують, що до 2025 року 70% ІТ-фахівців працюватимуть віддалено принаймні частково. Однак для досягнення стійкої ефективності необхідно балансувати між технологічними інноваціями, економічною вигодою та соціальним добробутом працівників.

1.2 Постановка задачі

Зростання популярності дистанційної роботи зумовлене технологічними змінами, пандемією COVID-19 та прагненням до більшої гнучкості робочих умов. Впровадження таких умов супроводжується значними викликами, серед яких:

- забезпечення продуктивної комунікації між працівниками;
- підтримка балансу між роботою та особистим життям;
- контроль за виконанням завдань і ефективністю співробітників;
- управління психологічним станом персоналу та уникнення ізоляції;
- забезпечення кібербезпеки та захисту даних при віддаленій роботі;
- створення ефективної системи мотивації віддалених працівників.

Ключовими факторами, що впливають на успішність дистанційної роботи, є наявність ефективних технологій для комунікації, чітко визначені ролі та обов'язки працівників, а також інструменти для моніторингу ефективності.

Основними викликами дистанційної роботи є проблеми зі спілкуванням, балансом роботи та особистого життя, контролем продуктивності та ментальним здоров'ям працівників (див. рис. 1.2).

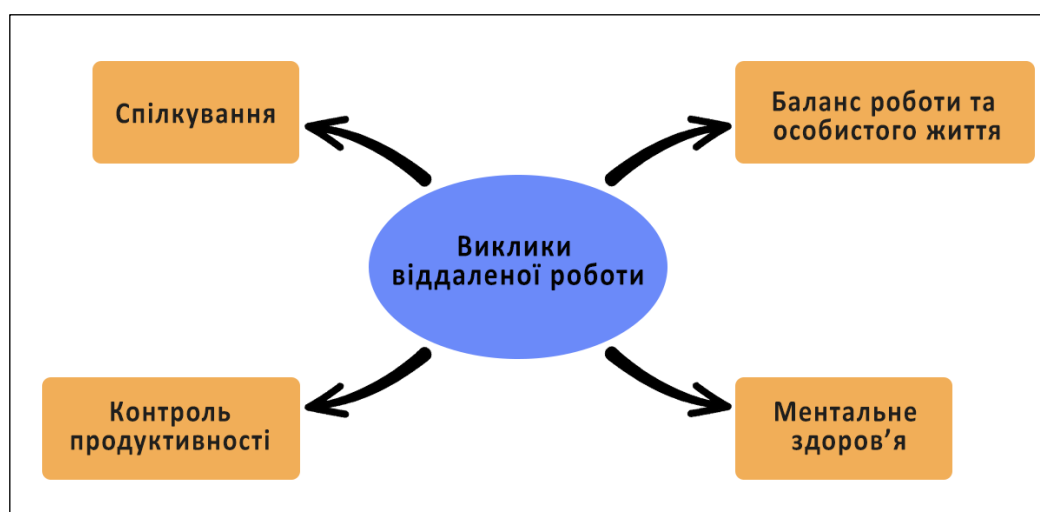


Рисунок 1.2 – Основні виклики дистанційної роботи (самостійне виконання)

Сучасна практика дистанційної зайнятості вимагає від ІТ-компаній гнучких і точних інструментів для оцінки ефективності персоналу, виявлення ризиків зниження продуктивності або звільнення, а також формування рекомендацій для підтримки балансу між роботою та добробутом працівників. Традиційні засоби управління персоналом не завжди відповідають цим вимогам, особливо в умовах гібридних або повністю віддалених форматів роботи.

У відповідь на ці виклики необхідно реалізувати дослідницький проєкт зі створення аналітичної системи для підтримки прийняття рішень на основі HR-даних. Для цього сформульовано низку практичних і дослідницьких завдань:

- аналіз факторів, що впливають на продуктивність працівників у дистанційному середовищі. За основу взято реальні структуровані дані з наборів `remote-work-productivity.csv` і IBM HR Analytics, а також проведено аналіз кореляцій між індикаторами навантаження, добробуту, рівня задоволеності й продуктивності;
- розширення набору ознак шляхом синтетичної генерації на основі існуючих даних. Оскільки деякі важливі характеристики були відсутні в початковому наборі, їх було додано шляхом моделювання залежностей між рівнем добробуту (`WellBeingScore`) та продуктивністю (`ProductivityScore`), із подальшим обчисленням `JobSatisfaction`, `EnvironmentSatisfaction` та `WorkLifeBalance`;
- побудова моделі прогнозування продуктивності працівника. Реалізовано регресійну модель на основі `Random Forest Regressor`, що дозволяє оцінити майбутню ефективність працівника на основі ключових факторів (робоче навантаження, тип зайнятості, вік, добробут, тощо);
- побудова моделі оцінки ймовірності звільнення (ризик плинності кадрів). Використано класифікаційний алгоритм `Random Forest Classifier` для прогнозування ризику звільнення працівника на основі його стану, умов праці, добробуту й інших показників;
- застосування кластеризації для аналізу типових груп працівників. Алгоритм `K-Means` дозволив згрупувати працівників за спільними

характеристиками, що дало змогу більш глибоко аналізувати специфіку їхньої поведінки та потреби;

- розробка інтерпретованої логіки формування рекомендацій. На основі кластеру, прогнозу продуктивності та рівня ризику формуються адаптивні рекомендації (наприклад, зниження навантаження, підвищення гнучкості, підтримка ментального здоров'я), які система надає у зрозумілому для HR-спеціаліста форматі;
- реалізація веб-застосунку на Flask. Кінцевий інтерфейс надає зручну форму для вводу даних, виводить прогнози, пояснення, візуалізації та пропозиції щодо управлінських рішень. Передбачено масштабування системи для подальшого використання у внутрішніх HR-аналітичних процесах.

Методологія дослідження базується на використанні:

- статистичного аналізу, масштабування ознак (StandardScaler);
- класифікаційного й регресійного моделювання (Random Forest);
- кластеризації (K-Means) для виявлення типових профілів працівників;
- оцінювання важливості ознак та побудови інтерпретаційних пояснень;
- інтеграції всіх моделей у веб-систему з зворотним зв'язком.

Наукова новизна дослідження полягає в комбінації синтетично збагаченого набору HR-даних, ансамблевих моделей машинного навчання та пояснювальних механізмів, що разом формують гнучкий інструмент для підтримки прийняття рішень у сфері управління дистанційними командами.

Практична значущість розробки полягає у можливості швидкого застосування прототипу в малих та середніх ІТ-компаніях без потреби у складній інфраструктурі, що особливо важливо в умовах динамічних змін ринку праці.

2 ОГЛЯД Й АНАЛІЗ ЛІТЕРАТУРНИХ ТА НАУКОВИХ ДЖЕРЕЛ

2.1 Аналіз глобальних наукових джерел

У сучасному світі інформаційних технологій дистанційна робота стала невід'ємною частиною професійного середовища. Пандемія COVID-19 значно прискорила цей перехід, і багато компаній були змушені адаптуватися до нових умов праці. Згідно з даними Бюро статистики праці США, у період з 2019 по 2021 роки відсоток дистанційних працівників зріс у всіх основних галузях, причому найбільше зростання спостерігалось в інформаційних технологіях [5].

Дослідження Стенфордського університету показало, що гібридна модель роботи, яка поєднує віддалену та офісну працю, не впливає на продуктивність працівників, але значно підвищує рівень їхньої задоволеності та знижує кількість звільнень [6]. Це свідчить про те, що гібридна форма організації праці може бути оптимальною для багатьох ІТ-компаній.

Однак, не всі аспекти дистанційної роботи є позитивними. Дослідження, опубліковане в журналі «Human Resource Management Journal», виявило, що працівники, які працюють віддалено, можуть відчувати підвищений рівень стресу та гніву порівняно з тими, хто працює в офісі [7]. Це підкреслює важливість підтримки психічного здоров'я працівників у віддаленому режимі.

Крім того, дослідження, опубліковане в «Journal of Business Research», підкреслює, що інноваційні організації, які впроваджують дистанційну роботу, повинні приділяти особливу увагу добробуту своїх працівників, оскільки це безпосередньо впливає на їхню продуктивність та залученість [8]. Незважаючи на виклики, багато компаній продовжують підтримувати гібридні моделі роботи.

Важливим аспектом є також фізичне здоров'я працівників. Збільшення часу, проведеного перед екраном, та малорухливий спосіб життя можуть призвести до проблем зі здоров'ям, таких як «техно-шиї» та інші м'язово-скелетні розлади [9]. Тому компаніям варто впроваджувати програми, спрямовані на підтримку фізичного здоров'я працівників, незалежно від їхнього місця роботи.

Загалом, дистанційна та гібридна робота стали невід'ємною частиною сучасної ІТ-галузі. Вони надають як можливості, так і виклики, які потребують

уважного аналізу та адаптації з боку компаній для забезпечення ефективності та добробуту своїх працівників.

2.2 Аналіз наукових джерел університету

В рамках університетських досліджень значна увага приділяється дослідженням у сфері штучного інтелекту, аналізу даних, інтелектуальних систем підтримки прийняття рішень та цифрової трансформації бізнес-процесів. Особливу увагу приділено створенню пояснюваних інтелектуальних систем, які дозволяють користувачам не лише отримувати прогнози, а й розуміти логіку, що стоїть за цими прогнозами.

Пояснення мають обґрунтовувати запропоновані інтелектуальною системою результати аналогічно людині – спеціалісту у відповідній предметній області [10].

Цей принцип безпосередньо реалізовано в аналітичній системі, розробленій у рамках кваліфікаційної роботи. Кожне числове прогнозне значення (продуктивність, ризик звільнення) супроводжується генерацією текстового пояснення, що враховує внесок окремих факторів, наприклад: «Працівник має високе навантаження та низький добробут, що може свідчити про вигорання». Це сприяє більш глибокому розумінню результатів як з боку HR-фахівців, так і керівників команд.

Дослідження в сфері психології пізнання показують, що людина зазвичай має отримати обґрунтування нових знань [11]. Цей підхід підтримує ідею, що без пояснень навіть точні прогнози значення можуть бути проігноровані користувачем або викликати недовіру. Врахування цього аспекту стало мотивацією для включення механізму генерації текстових рекомендацій у систему.

Крім теоретичних публікацій, результати подібних підходів знаходять відображення у попередніх магістерських роботах студентів ХНУРЕ. Так, у дослідженнях останніх років розглядалися системи підтримки прийняття рішень у медичній діагностиці, емоційному аналізі, автоматизованому консультуванні та HR-аналітиці. Накопичений досвід дозволив побудувати інтегровану аналітичну

систему, яка поєднує кластеризацію, регресійне прогнозування та інтерпретовані моделі машинного навчання.

Таким чином, концептуальні підходи, розроблені у межах наукової школи ХНУРЕ, безпосередньо вплинули на архітектуру створеного інструменту та методологію проведеного дослідження. Це дозволило забезпечити не лише технічну точність, але й зручність використання та довіру до результатів з боку кінцевого користувача.

2.3 Узагальнення підходів до прогнозування продуктивності та ризику звільнення

На основі проаналізованих наукових джерел можна виділити два основні напрями сучасних досліджень, пов'язаних з ефективністю працівників у дистанційному форматі: прогнозування рівня продуктивності та оцінювання ризику звільнення. Обидва напрями тісно пов'язані між собою та зумовлені спільними викликами – обмеженим доступом до традиційних показників ефективності, розмитими часовими межами роботи та зростанням значущості суб'єктивних факторів, таких як добробут і задоволеність.

У сфері прогнозування продуктивності дослідники використовують як прямі, так і непрямі показники. До прямих належать обсяг виконаних завдань, дотримання дедлайнів, кількість комунікацій тощо. Непрямі включають індикатори добробуту, задоволеності умовами праці, балансу між роботою та особистим життям. Значну увагу приділяється побудові регресійних моделей, що враховують як кількісні (кількість годин, заробітна плата, вік), так і якісні (тип зайнятості, емоційний стан) змінні.

Щодо ризику звільнення, актуальними є класифікаційні моделі, які дозволяють розподіляти працівників за ймовірністю залишити компанію. У цьому контексті поширеними є моделі на основі рішень дерев (зокрема Random Forest), що забезпечують інтерпретованість і високу точність при роботі з невеликими, але багатовимірними вибірками. Основними предикторами виступають такі ознаки, як

участь в понаднормовій роботі, рівень задоволеності роботою та зміни заробітної плати.

Узагальнюючи, можна стверджувати, що ефективне прогнозування продуктивності та ризику звільнення можливе лише при поєднанні об'єктивних метрик із суб'єктивними оцінками. Саме цей підхід дозволяє врахувати індивідуальні особливості працівника, зокрема у віддаленому форматі роботи, де традиційні KPI втрачають частину своєї інформативності. Водночас зростає роль аналітичних моделей з механізмами пояснюваності (XAI), які забезпечують прозорість рішень та підтримку з боку HR-фахівців.

Таким чином, огляд наукових джерел підтверджує доцільність поєднання регресійного, класифікаційного та кластерного аналізу як базових підходів до розв'язання поставленої задачі.

3 ПРОЄКТНІ РІШЕННЯ І АРХІТЕКТУРА СИСТЕМИ

3.1 Загальна архітектура програмної системи

У межах дослідження було розроблено аналітичну інформаційну систему, призначену для індивідуального аналізу ефективності працівників, що працюють у дистанційному або гібридному форматі. Система дозволяє прогнозувати рівень продуктивності, оцінювати ризик звільнення та формувати персоналізовані управлінські рекомендації для HR-фахівців.

Загальна архітектура реалізована за принципами клієнт-серверної взаємодії (див. рис. 3.1) та включає наступні логічні компоненти:

- інтерфейс користувача (Frontend) – реалізований за допомогою HTML, CSS і JavaScript. Представляє собою вебсторінку з адаптивною формою введення параметрів працівника. Після надсилання форми система генерує відповідь, що включає прогнозовану продуктивність, ризик звільнення, тип працівника та текстову рекомендацію;
- серверна частина (Backend) – побудована з використанням Python-фреймворку Flask. Вона виконує обробку вхідних запитів, попередню обробку даних, звернення до моделей машинного навчання та формування результатів у форматі JSON;
- аналітичне ядро – складається з попередньо натренованих моделей машинного навчання (класифікація, регресія, кластеризація), що збережені у вигляді серіалізованих об'єктів (.pkl). Їх завантаження й виклик відбувається на боці сервера.

Передача даних між клієнтською та серверною частинами реалізується через HTTP POST-запити. На сервері запит обробляється у спеціальному ендпоінті /predict, що приймає структуру JSON, проводить масштабування даних, надсилає їх у відповідні моделі та повертає результати аналізу.

Архітектура є модульною та масштабованою, що забезпечує можливість інтеграції додаткових функцій, таких як груповий аналіз або збереження історії запитів. Обчислювальні ресурси оптимізовано через використання легковагового

фреймворку Flask, що дозволяє розгортання як на локальних серверах, так і в хмарному середовищі.

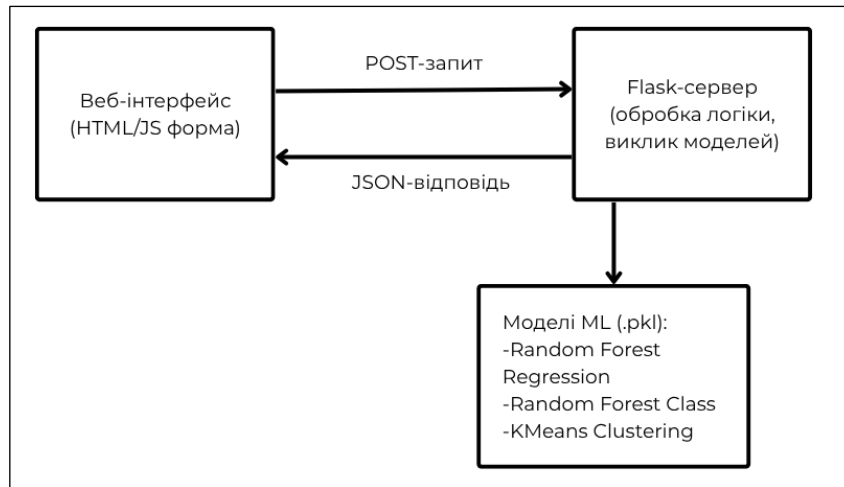


Рисунок 3.1 – Загальна архітектура аналітичної системи (самостійне виконання)

3.2 Структура веб-застосунку

Веб-застосунок, розроблений у межах кваліфікаційної роботи, реалізує архітектуру типу «клієнт-сервер» та складається з трьох основних підсистем: інтерфейсу користувача (frontend), серверної логіки обробки запитів (backend) і модуля аналітики на основі моделей машинного навчання.

Інтерфейс користувача (frontend) реалізовано засобами HTML, CSS та JavaScript. Він забезпечує взаємодію HR-користувача із системою через веб-форму, в якій вводяться характеристики працівника. Після заповнення даних здійснюється формування JSON-запиту, що надсилається на сервер асинхронно. Основні файли, що забезпечують функціонування клієнтської частини, розміщені у директоріях `templates/` (файл `index.html`) та `static/` (файл `script.js`).

Серверна частина (backend) реалізована з використанням Python-фреймворку Flask. Основний програмний модуль – `app.py`, у якому відбувається ініціалізація вебсервера, обробка HTTP-запитів, а також зв'язок із аналітичним ядром. Сервер приймає запити через маршрут `/predict`, виконує валідацію даних, передобробку (за допомогою модуля `utils/preprocessing.py`) та надсилає вхідні значення до відповідних моделей. Після отримання результатів здійснюється генерація

персоналізованих рекомендацій (модуль `utils/recommend.py`), які разом із прогнозами формуються у відповідь.

Аналітичне ядро включає набір серіалізованих моделей машинного навчання, збережених у директорії `model/`. Система використовує:

- `productivity_model.pkl` – регресійну модель для прогнозування ефективності працівника;
- `attrition_model.pkl` – класифікатор для визначення ризику звільнення;
- `cluster_model.pkl` – модель кластеризації для типізації працівників;
- `scaler.pkl` та `scaler_attrition.pkl` – масштабувальники, навчені на відповідних підмножинах даних;
- `feature_importance.csv` та `productivity_feature_importance.csv` – допоміжні артефакти, що містять оцінки значущості ознак.

Файлова структура наведена нижче (див. рис. 3.2). Вона чітко поділена на логічні блоки, що відповідають за взаємодію з користувачем, обробку запитів, машинне навчання, візуалізацію та обслуговування API.

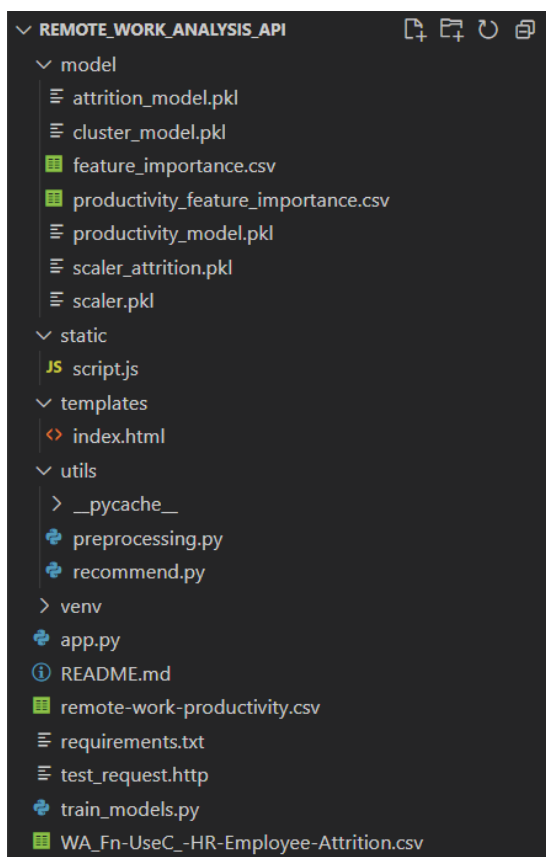


Рисунок 3.2 – Структура файлів і модулів веб-застосунку (самостійне виконання)

Крім основного функціоналу, у корені проєкту також розміщено навчальні набори даних (`remote-work-productivity.csv`, `WA_Fn-UseC_-HR-Employee-Attrition.csv`), файл `test_request.http` для тестування API у форматі HTTP, конфігураційний файл залежностей `requirements.txt`, а також інструкцію з використання `README.md`. Локальне середовище ізольовано у директорії `venv/`.

Розроблена структура забезпечує чіткий розподіл відповідальностей між компонентами, що сприяє підтримуваності системи, зручності її розгортання та подальшої інтеграції у внутрішні процеси організацій.

3.3 Взаємодія компонентів та схема використання моделей

У межах створеної аналітичної системи взаємодія її складових елементів реалізована за принципом наскрізного потоку обробки даних. Архітектура забезпечує узгоджену роботу між клієнтською частиною, серверною логікою та аналітичним ядром. Такий підхід дозволяє реалізувати ефективне прогнозування продуктивності працівника, виявлення ризику його звільнення, а також формування релевантних рекомендацій.

Першим етапом взаємодії є введення даних користувачем через вебінтерфейс. Інтерфейс розроблено у вигляді HTML-форми з адаптивною версткою. Користувач (HR-фахівець або менеджер) вводить значення ключових параметрів: кількість годин, відпрацьованих за тиждень, рівень добробуту, тип зайнятості, вік, заробітну плату та інші показники. На основі цього формується асинхронний HTTP-запит типу POST.

Отримані дані надсилаються на сервер, що реалізований за допомогою фреймворку Flask. На сервері запит надходить до ендпоінту `/analyze`, який виконує валідацію, перевірку повноти та первинну обробку даних. Цей процес реалізовано у модулі `preprocessing.py`. Масштабування вхідних числових ознак виконується за допомогою збережених об'єктів `StandardScaler`, навчення яких здійснювалося під час побудови моделей.

Після підготовки дані передаються до трьох окремих моделей машинного навчання, що зберігаються у серіалізованому вигляді у форматі .pkl. Кожна з трьох моделей має своє призначення:

- регресійна модель (productivity_model.pkl) оцінює рівень продуктивності працівника на шкалі від 0 до 100;
- класифікаційна модель (attrition_model.pkl) визначає ризик звільнення, подаючи його у вигляді ймовірності;
- модель кластеризації (cluster_model.pkl) відносить працівника до одного з кластерів – «стабільний», «перевантажений» або «з ознаками вигорання».

Результати прогнозування передаються до модуля формування рекомендацій (recommend.py). Цей модуль містить логіку створення текстових висновків для HR-менеджера з урахуванням поєднання продуктивності, ризику звільнення та типу працівника. Наприклад, якщо виявлено низький добробут і високий ризик звільнення – система запропонує надати працівнику відпустку або зменшити навантаження.

Формалізована відповідь генерується у форматі JSON і надсилається назад на клієнтську частину. Інтерфейс оновлюється динамічно, виводячи числові результати, класифікацію працівника та супровідну рекомендацію.

Таким чином, реалізовано повноцінний цикл обробки запиту (див. рис. 3.3).

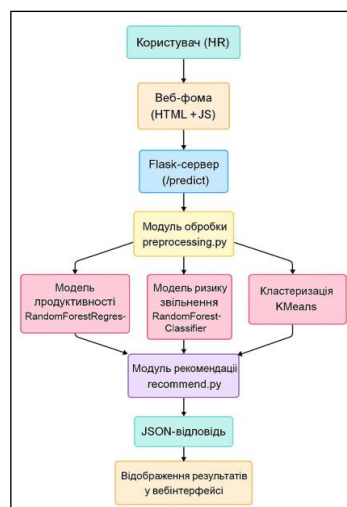


Рисунок 3.3 – Послідовність етапів взаємодії між компонентами системи (самостійне виконання)

Він включає збір даних, підготовку ознак, аналітичну обробку, інтерпретацію результатів і генерацію відповідей у зрозумілому для користувача форматі. На діаграмі зображено логічну послідовність етапів взаємодії між компонентами системи.

4 ОПИС ПРОГРАМНОЇ РЕАЛІЗАЦІЇ ТА МЕТОДІВ МАШИННОГО НАВЧАННЯ

4.1 Підготовка навчальних даних та обробка ознак

Процес машинного навчання є чутливим до якості вхідних даних. Тому перед побудовою моделей було здійснено ретельну підготовку датасетів, у тому числі аналіз доступних ознак, нормалізацію значень, формування нових ознак та заповнення відсутніх значень. У межах цієї роботи було використано два незалежні відкриті набори даних (Remote Work Productivity Dataset [12] та IBM HR Analytics Dataset [13]), кожен із яких призначений для розв'язання окремої задачі: регресії (оцінка продуктивності працівника) та класифікації (оцінка ризику звільнення) та кластеризації (типізація працівників за схожими ознаками). Їх комбіноване використання дозволило сформувати більш комплексну основу для побудови та перевірки моделей. Характеристики обох датасетів наведено нижче (див. табл. 4.1).

Таблиця 4.1 – Опис застосованих наборів даних (таблиця виконана самостійно)

Назва датасету	Джерело	Кількість записів	Ключові поля	Призначення записів
Remote Work Productivity	Kaggle	1000	HoursWorkedPerWeek, WellBeingScore, EmploymentType, ProductivityScore	Побудова моделі регресії та кластеризації
IBM HR Analytics	IBM, Kaggle	1000	Age, MonthlyIncome, JobSatisfaction, Attrition, OverTime, WorkLifeBalance	Побудова класифікаційної моделі

Кожен з наборів використовувався окремо. Об'єднання датасетів не здійснювалося. Це дозволило зберегти цілісність кожної вибірки та уникнути спотворень у розподілах.

У датасеті Remote Work Productivity були відсутні такі ознаки, як JobSatisfaction, EnvironmentSatisfaction, WorkLifeBalance, які є типовими для HR-аналітики. З метою їх відтворення було здійснено генерацію на основі двох

ключових параметрів — добробуту (*WellBeingScore*) та оцінки продуктивності (*ProductivityScore*). Формула для генерації мала вигляд:

$$score = round\left(\frac{(0.6 \cdot WellBeingScore + 0.4 \cdot ProductivityScore)}{100} \cdot 3 + 1 + \varepsilon\right) \quad (4.1)$$

де ε – випадкова величина з нормальним розподілом ($\varepsilon \sim N(0,0.3)$). Згенеровані значення обмежувалися в межах $[1, 4]$, що відповідає шкалам оцінювання задоволеності в IBM HR Dataset.

Для забезпечення стабільної роботи моделей машинного навчання всі числові характеристики було стандартизовано за допомогою *StandardScaler*, з використанням формули:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (4.2)$$

де x_i – значення ознаки,

μ – середнє значення по вибірці,

σ – стандартне відхилення.

Цей підхід дозволив уникнути зміщення моделей у бік ознак з великою дисперсією, таких як заробітна плата або вік.

Моделі машинного навчання працювали із трьома задачами, відповідно до яких і формувались цільові змінні:

- модель регресії (*RandomForestRegressor*) – прогнозує *ProductivityScore* на основі ознак з *Remote Work Productivity*;
- модель класифікації (*RandomForestClassifier*) – передбачає значення *Attrition* (0 або 1) за даними *IBM HR Analytics*;
- модель кластеризації (*KMeans*) – формує групи працівників без використання цільової змінної, лише за схожістю вхідних параметрів.

Для обох моделей із учителем (регресія та класифікація) було застосовано перехресну валідацію із використанням GridSearchCV (параметр cv=5). Це дозволило автоматично підбирати оптимальні гіперпараметри моделі та зменшити ризик переобучення. Поділ на тренувальну й тестову вибірки явно не виконувався, оскільки функція GridSearchCV реалізує внутрішню стратегію оцінки на незалежних підвбірках.

4.2 Вибір моделей та обґрунтування алгоритмів

Вибір моделей машинного навчання здійснювався з урахуванням характеру задачі, структури доступних даних, а також необхідності отримання інтерпретованих та стійких результатів для подальшого використання в HR-аналітиці. У межах дослідження реалізовано три незалежні моделі: регресійну, класифікаційну та кластеризаційну. Алгоритми обрано за принципами точності, стабільності та інтерпретованості.

Для задачі прогнозування оцінки продуктивності (ProductivityScore) було обрано алгоритм Random Forest Regressor.

Алгоритм належить до ансамблевих методів, заснованих на ідеї побудови колекції дерев рішень і усереднення їх результатів.

Переваги алгоритму:

- стійкість до перенавчання;
- автоматичне врахування взаємодій між ознаками;
- здатність працювати з даними різної природи без масштабування;
- збереження важливості ознак (feature importance) для пояснення моделі.

Цей підхід дозволив досягти високої точності при мінімальному налаштуванні гіперпараметрів, а також забезпечити стабільність моделі на малих обсягах даних.

Для оцінки ризику звільнення працівника (Attrition) було також використано алгоритм Random Forest Classifier. Це класифікаційний варіант попередньої моделі, де кожне дерево голосує за певний клас, а фінальне рішення ухвалюється більшістю голосів.

Обґрунтування вибору:

- хороша робота з розбалансованими вибірками;
- здатність виявляти нетривіальні залежності у HR-даних;
- високі метрики якості (accuracy, precision, recall) на тестових розбиттях у попередніх дослідженнях;
- інтерпретованість при подальшій візуалізації результатів (наприклад, за допомогою дерев або SHAP-графіків).

Алгоритм також добре адаптується до категоріальних ознак і дозволяє швидко оцінити вплив окремих характеристик працівника на ймовірність його звільнення.

Для сегментації працівників за поведінковими та соціальними характеристиками використано алгоритм K-Means, який дозволяє групувати спостереження на основі мінімізації внутрішньогрупового розсіювання:

Обґрунтування вибору:

- простота реалізації та швидкість обчислень;
- ефективність на великих об'ємах нормалізованих числових даних;
- можливість виявлення прихованих типів працівників (стабільні, перевантажені, схильні до вигорання);
- використання результатів кластеризації для генерації персоналізованих рекомендацій у системі.

Кількість кластерів $k=3$ було обрано емпірично шляхом попереднього аналізу інерції та візуалізації даних.

4.3 Інтерпретація результатів та формування рекомендацій

Результати роботи моделей машинного навчання у дослідженні дозволили виявити ключові залежності між поведінковими, психологічними та продуктивними характеристиками працівників. Це дало змогу побудувати персоналізовану систему рекомендацій та забезпечити інтерпретованість для HR-фахівців.

Модель RandomForestClassifier, навчена на датасеті IBM HR Analytics, дозволяє прогнозувати ризик звільнення на підставі таких параметрів, як задоволеність, дохід, вік і наявність понаднормової роботи (див. табл. 4.2).

Таблиця 4.2 – Вагомість ознак у класифікаційній моделі (таблиця виконана самостійно)

Ознака	Вагомість (імпакт)
OverTime	0.319
JobSatisfaction	0.218
MonthlyIncome	0.174
EnvironmentSatisfaction	0.138
Age	0.088
WorkLifeBalance	0.063

Результати підтверджують: наявність понаднормової роботи (OverTime) є головним предиктором звільнення, а також важливу роль відіграють психологічні фактори (JobSatisfaction, EnvironmentSatisfaction).

Модель RandomForestRegressor оцінює продуктивність працівника за шкалою від 0 до 100 балів. Аналіз результатів показав наступні закономірності:

- високий рівень добробуту корелює з більшою продуктивністю;
- оптимальне навантаження — 38–45 годин на тиждень;
- низький рівень WorkLifeBalance зазвичай призводить до нижчих результатів.

Таблиця 4.3 – Добробут та прогнозована продуктивність (таблиця виконана самостійно)

WellBeingScore (діапазон)	Середня продуктивність
20-40	58-63
40-60	65-71
60-80	72-77
80-100	78-85

Алгоритм KMeans дозволив здійснити сегментацію працівників на три кластери, кожен з яких відповідає певному поведінковому профілю. Це дозволяє не

лише виявляти проблеми, але й розробляти диференційовані стратегії підтримки персоналу.

Таблиця 4.4 – Типи працівників за результатами кластеризації (таблиця виконана самостійно)

Кластер	Тип працівника	Коротка характеристика
0	Стабільний	Збалансоване навантаження, середній добробут, низький ризик
1	Перевантажений	Високі години роботи, продуктивність вище середньої, ризик вигорання
2	Уразливий	Низький добробут, незадоволеність, високий ризик звільнення

Система формує текстові рекомендації для HR-фахівців з урахуванням:

- кластеру працівника;
- оцінки продуктивності;
- ризику звільнення;
- добробуту та задоволеності.

Приклад згенерованої рекомендації: «Працівник демонструє ознаки перевантаження. Незважаючи на високу продуктивність (81/100), ризик звільнення становить 41%. Рекомендовано обмежити понаднормові години, запропонувати додаткові дні відпочинку.»

Попри високу точність, модель не враховує:

- динаміку поведінки з часом;
- соціально-психологічні аспекти (атмосфера в команді, мотивація);
- зовнішні фактори (економічні кризи, індивідуальні життєві обставини).

5 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

5.1 Характеристика вхідних даних та підготовка до тестування

З метою перевірки працездатності розробленої аналітичної системи було проведено серію експериментальних запусків із тестовими прикладами, які імітують типові HR-сценарії. Основна мета цього етапу дослідження – оцінити, наскільки система здатна коректно обробляти введені параметри працівника, виконувати аналітичну обробку, прогнозувати рівень продуктивності, виявляти ризик звільнення та формувати персоналізовані рекомендації.

Вхідні дані можуть подаватися у двох форматах:

- через веб-інтерфейс у вигляді форми введення;
- програмно через API-запит у форматі JSON-об'єкта.

Набір вхідних ознак охоплює ключові характеристики, що впливають на ефективність працівника, його навантаження, загальний добробут та ризики емоційного вигорання. Нижче наведено перелік основних параметрів, які використовуються у системі:

- `HoursWorkedPerWeek` – середнє тижневе навантаження (кількість годин);
- `WellBeingScore` – інтегральна оцінка добробуту працівника (від 1 до 100);
- `JobSatisfaction` – рівень задоволеності роботою (1–4);
- `WorkLifeBalance` – власна оцінка балансу між роботою та особистим життям працівника (1–4);
- `EnvironmentSatisfaction` – комфортність умов праці (1–4);
- `EmploymentType` – тип зайнятості (дистанційна або офісна);
- `OverTime` – факт наявності понаднормової роботи (так/ні);
- `Age` – вік працівника;
- `MonthlyIncome` – рівень заробітної плати (у доларах США).

Для моделей машинного навчання усі дані проходять попередню обробку. Числові ознаки масштабуються за допомогою `StandardScaler`, категоріальні – кодуються у відповідному форматі: one-hot або бінарному. На основі оброблених даних формується єдиний вектор ознак, який передається до трьох моделей:

регресійної (продуктивність), класифікаційної (ризик звільнення) та кластеризаційної (тип працівника).

Задля демонстрації функціонування системи було сформовано низку тестових записів, які імітують реальні ситуації. Один із прикладів наведено у таблиці:

Таблиця 5.1 – Приклад вхідних параметрів для тестування (таблиця виконана самостійно)

Ознака	Значення
Вік	34
Тип зайнятості	Дистанційна
Рівень добробуту	72
Задоволеність роботою	3
Баланс між роботою та життям	3
Задоволеність умовами	3
Кількість годин роботи на тиждень	46
Чи працює понаднормово	Так
Місячний дохід (в доларах)	8700

Для більш повного розуміння структури використаних даних на рисунках нижче подано фрагменти двох основних датасетів, які застосовувалися під час навчання моделей. Фрагмент датасету IBM HR Analytics Dataset наведено нижче (див. рис. 5.1).

```

WA_Fn-UseC-HR-Employee-Attrition.csv
1 Age,Attrition,BusinessTravel,DailyRate,Department,DistanceFromHome,Education,EducationField,EmployeeCount,EmployeeNumber,EnvironmentSatisfact
2 41,Yes,Travel_Rarely,1102,Sales,1,2,Life_Sciences,1,1,2,Female,94,3,2,Sales_Executive,4,Single,5993,19479,8,Y,Yes,11,3,1,80,0,8,0,1,6,4,0,5
3 49,No,Travel_Frequently,279,Research & Development,8,1,Life_Sciences,1,2,3,Male,61,2,2,Research_Scientist,2,Married,5130,24907,1,Y,No,23,4,4,
4 37,Yes,Travel_Rarely,1373,Research & Development,2,2,Other,1,4,4,Male,92,2,1,Laboratory_Technician,3,Single,2090,2396,6,Y,Yes,15,3,2,80,0,7,3
5 33,No,Travel_Frequently,1392,Research & Development,3,4,Life_Sciences,1,5,4,Female,56,3,1,Research_Scientist,3,Married,2909,23159,1,Y,Yes,11,
6 27,No,Travel_Rarely,591,Research & Development,2,1,Medical,1,7,1,Male,40,3,1,Laboratory_Technician,2,Married,3468,16632,9,Y,No,12,3,4,80,1,6,
7 32,No,Travel_Frequently,1005,Research & Development,2,2,Life_Sciences,1,8,4,Male,79,3,1,Laboratory_Technician,4,Single,3068,11864,0,Y,No,13,3
8 59,No,Travel_Rarely,1324,Research & Development,3,3,Medical,1,10,3,Female,81,4,1,Laboratory_Technician,1,Married,2670,9964,4,Y,Yes,20,4,1,80,
9 30,No,Travel_Rarely,1358,Research & Development,24,1,Life_Sciences,1,11,4,Male,67,3,1,Laboratory_Technician,3,Divorced,2693,13335,1,Y,No,22,4
10 38,No,Travel_Frequently,216,Research & Development,23,3,Life_Sciences,1,12,4,Male,44,2,3,Manufacturing_Director,3,Single,9526,8787,0,Y,No,21,
11 36,No,Travel_Rarely,1299,Research & Development,27,3,Medical,1,13,3,Male,94,3,2,Healthcare_Representative,3,Married,5237,16577,6,Y,No,13,3,2,
12 35,No,Travel_Rarely,809,Research & Development,16,3,Medical,1,14,1,Male,84,4,1,Laboratory_Technician,2,Married,2426,16479,0,Y,No,13,3,3,80,1,
13 29,No,Travel_Rarely,453,Research & Development,15,2,Life_Sciences,1,15,4,Female,49,2,2,Laboratory_Technician,3,Single,4193,12682,0,Y,Yes,12,3
14 31,No,Travel_Rarely,670,Research & Development,26,1,Life_Sciences,1,16,1,Male,31,3,1,Research_Scientist,3,Divorced,2911,15170,1,Y,No,17,3,4,8
15 34,No,Travel_Rarely,1346,Research & Development,19,2,Medical,1,18,2,Male,93,3,1,Laboratory_Technician,4,Divorced,2661,8758,0,Y,No,11,3,3,80,1
16 28,Yes,Travel_Rarely,103,Research & Development,24,3,Life_Sciences,1,19,3,Male,50,2,1,Laboratory_Technician,3,Single,2028,12947,5,Y,Yes,14,3,
17 29,No,Travel_Rarely,1389,Research & Development,21,4,Life_Sciences,1,20,2,Female,51,4,3,Manufacturing_Director,1,Divorced,9980,10195,1,Y,No,1
18 32,No,Travel_Rarely,334,Research & Development,5,Col 7: Education,21,1,Male,80,4,1,Research_Scientist,2,Divorced,3298,15053,0,Y,Yes,12,3,4,8
19 22,No,Non-Travel,1123,Research & Development,16,2,Medical,1,22,4,Male,96,4,1,Laboratory_Technician,4,Divorced,2935,7324,1,Y,Yes,13,3,2,80,2,1

```

Рисунок 5.1 – Фрагмент датасету IBM HR Analytics (самостійне виконання)

Фрагмент датасету Remote Work Productivity Dataset наведено нижче (див. рис. 5.2)

```

remote-work-productivity.csv
1 EmployeeID;EmploymentType;HoursWorkedPerWeek;ProductivityScore;WellBeingScore
2 1000;Remote;34;80;82
3 999;In-Office;57;50;52
4 998;In-Office;45;74;61
5 997;Remote;33;88;73
6 996;Remote;33;88;82
7 995;In-Office;40;60;56
8 994;In-Office;55;62;25
9 993;In-Office;41;71;80
10 992;In-Office;36;73;76
11 991;In-Office;57;49;30
12 990;In-Office;49;44;56
13 989;In-Office;61;31;37
14 988;In-Office;52;53;49
15 987;In-Office;45;65;53
16 986;In-Office;44;63;57

```

Рисунок 5.2 – Фрагмент датасету Remote Work Productivity (самостійне виконання)

Кожен набір даних, поданий у систему, проходить наступні етапи:

- валідація параметрів: перевірка на повноту та допустимі значення;
- масштабування числових ознак: використовується серіалізований об'єкт StandardScaler, який було збережено під час тренування моделі;
- кодування категоріальних змінних: ознаки типу зайнятості та понаднормової роботи перетворюються у формат one-hot або бінарний;
- формування єдиного вектору ознак: для подання у всі три моделі.

Після цього запит передається до серверної частини, де виконується прогнозування.

Формат вхідних даних було узгоджено з типовими анкетами, які використовуються у внутрішніх HR-опитуваннях. Це забезпечує зручність застосування системи в умовах реального бізнес-середовища.

5.2 Аналіз результатів та приклади HR-сценаріїв

Після подачі вхідних даних аналітична система проводить автоматизоване прогнозування, класифікацію та генерацію HR-рекомендацій. Користувач вводить

ключові параметри працівника (навантаження, добробут, вік, тип зайнятості тощо), після чого система обробляє запит і відображає результати в інтерфейсі.

У процесі обчислень задіяно три окремі моделі:

- регресійна модель оцінює ймовірну продуктивність працівника (ProductivityScore) на шкалі від 0 до 100;
- класифікаційна модель визначає ризик звільнення (AttritionRisk) у відсотковому вираженні;
- кластеризатор відносить працівника до одного з трьох кластерів (0 – стабільний, 1 – перевантажений, 2 – ознаки вигорання).

У веб-інтерфейсі реалізовано інтуїтивно зрозумілу форму виведення результатів. Користувач бачить як числові значення продуктивності та ризику, так і текстову рекомендацію, що згенерована системою на основі кластеру та логіки ухвалення рішень (див. рис. 5.3).

The screenshot shows a web interface titled "HR Аналіз працівника". It contains several input fields and dropdown menus arranged in a grid:

- Кількість годин роботи на тиждень:** Input field with value 40.
- Рівень добробуту (0 = дуже низький, 100 = відмінний):** Input field with value 60.
- Задоволеність роботою:** Dropdown menu with value 3 — Висока.
- Задоволеність робочим середовищем:** Dropdown menu with value 3 — Висока.
- Баланс між роботою та життям:** Dropdown menu with value 4 — Відмінний.
- Формат роботи:** Dropdown menu with value В офісі.
- Вік (років):** Input field with value 33.
- Місячний дохід (\$):** Input field with value 5500.
- Чи працює понаднормово?:** Dropdown menu with value Так.

At the bottom left, there is a blue button labeled "Аналізувати".

Рисунок 5.3 – Веб-інтерфейс системи (самостійне виконання)

Після натискання кнопки «Аналізувати» користувач отримує структурований результат у форматі числових значень та текстового коментаря. Така форма виведення є максимально придатною для оперативного прийняття рішень у сфері HR (див. рис. 5.4).

<p>Результати</p> <p>Продуктивність: 71.9/100 — середня</p> <p>Ризик звільнення: 47% — середній ризик</p> <p>Пояснення: Прогнозована продуктивність: 71.9 балів — середня. Ймовірність звільнення: 0.47 — помірний ризик. Цей працівник демонструє збалансовану поведінку — середній рівень навантаження, продуктивності та добробуту. Ситуація виглядає стабільною, але варто періодично перевіряти рівень мотивації.</p> <p>Рекомендації:</p> <p>Продуктивність працівника знаходиться в прийнятних межах. Варто підтримувати наявні умови праці, щоб зберегти стабільний рівень ефективності.</p> <p>Існує помірний ризик звільнення. HR варто звернути увагу на добробут працівника, рівень задоволеності та рівень навантаження. Профілактична розмова може запобігти небажаному переходу.</p> <p>Цей працівник демонструє збалансовану поведінку — середній рівень навантаження, продуктивності та добробуту. Ситуація виглядає стабільною, але варто періодично перевіряти рівень мотивації.</p>
--

Рисунок 5.4 – Згенерована рекомендація (самостійне виконання)

Рекомендації, сформовані системою, є персоналізованими – вони враховують не лише значення продуктивності та ризику, а й тип поведінкового кластера, до якого належить працівник. Наприклад: «Працівник демонструє ознаки перевантаження. Незважаючи на високу продуктивність, ризик звільнення перевищує середній. Рекомендовано тимчасово зменшити обсяг задач або переглянути графік.»

Обробка запиту реалізована на сервері, де послідовно викликаються три серіалізовані моделі машинного навчання. Фрагмент відповідного серверного коду зображено нижче (див. рис. 5.5).

Цей код демонструє, як система обробляє запит, застосовує машинні моделі та генерує відповідь у форматі JSON. Серверна частина пов'язана з фронтендом через API, що забезпечує гнучкість інтеграції з іншими корпоративними сервісами. Такий підхід дозволяє масштабувати систему на більші обсяги даних або підключати її до внутрішніх HR-систем компаній.

```

# === Обробка форми аналізу працівника ===
@app.route('/analyze', methods=['POST'])
def analyze():
    data = request.json

    try:
        # Попередня обробка
        X_scaled, attrition_scaled = preprocess_input(data, scaler, scaler_attrition)

        # Прогнози
        cluster = int(cluster_model.predict(X_scaled)[0])
        predicted_productivity = float(productivity_model.predict(X_scaled)[0])
        attrition_prob = float(attrition_model.predict_proba(attrition_scaled)[0][1])

        # Отримання рекомендацій
        recommendations, explanation = get_recommendations(
            cluster=cluster,
            attrition_prob=attrition_prob,
            productivity=predicted_productivity
        )

        return jsonify({
            "cluster": cluster,
            "predicted_productivity": round(predicted_productivity, 1),
            "attrition_risk": round(attrition_prob, 2),
            "recommendations": recommendations,
            "explanation": explanation
        })

    except Exception as e:
        return jsonify({"error": str(e)}), 500

```

Рисунок 5.5 – Обробка запиту на сервері (самостійне виконання)

Розглянемо типові ситуації, в яких система може бути використана на практиці. Наведені нижче приклади демонструють гнучкість та адаптивність моделі до різних профілів працівників.

Сценарій 1 – ознаки вигорання: працівник, що працює понад 45 годин на тиждень, має середній рівень добробуту (WellBeingScore = 58) та низький рівень задоволеності роботою (2 з 4), класифікується системою як такий, що демонструє ознаки перевантаження. Система оцінює ризик звільнення на рівні 47%, а кластеризація відносить його до кластеру 2 (вигорання).

HR-рекомендація: провести індивідуальну зустріч, запропонувати скорочення обсягу завдань, організувати підтримку ментального здоров'я (внутрішній психолог, день без зустрічей тощо).

Сценарій 2 – високий ризик за низької продуктивності: у випадку працівника з низьким WellBeingScore (40), незадовільним WorkLifeBalance (1 з 4) та регулярною понаднормовою роботою, система фіксує низьку продуктивність (61/100) та високий ризик звільнення (62%). Такий працівник входить до кластеру 1 або 2 в залежності від інших показників, однак система сигналізує про критичність ситуації.

HR-рекомендація: ротація або зміна команди, пропозиція відпустки, направлення на консультацію до фахівця. У разі підтвердження вигорання – гнучкий графік або тимчасове зниження навантаження.

Сценарій 3 – стабільний працівник: працівник 39 років, працює 38 годин на тиждень, задоволений роботою (оцінка 4), має високий індекс добробуту (WellBeingScore = 88). Прогнозована продуктивність – 83/100, ризик звільнення – 12%. Кластеризація відносить його до кластеру 0 – стабільні працівники.

HR-рекомендація: заохочення за добрі результати, можливість просування по кар'єрних сходах або участь у внутрішніх програмах наставництва.

Для глибшого аналізу та підтримки прийняття стратегічних HR-рішень у системі реалізовано візуальні аналітичні компоненти. Одним із ключових є графік розподілу прогнозованої продуктивності персоналу – ProductivityScore, що будується автоматично після обробки вхідних даних усіх працівників або тестових профілів (див. рис. 5.6).

Такий графік дає змогу не лише візуально оцінити загальний стан продуктивності, а й виявити аномальні або граничні значення, які потребують додаткової уваги. Аналітика продуктивності за допомогою гістограми – це перевірений інструмент у сфері people analytics, оскільки він дозволяє швидко визначати розподіл зусиль та результатів серед команди.

Аналіз гістограми демонструє, що основна маса значень продуктивності знаходиться в діапазоні від 60 до 80 пунктів, що вказує на наближеність до нормального (гаусового) розподілу. Це підтверджує, що більшість працівників демонструють середній рівень ефективності, характерний для стабільної організації.

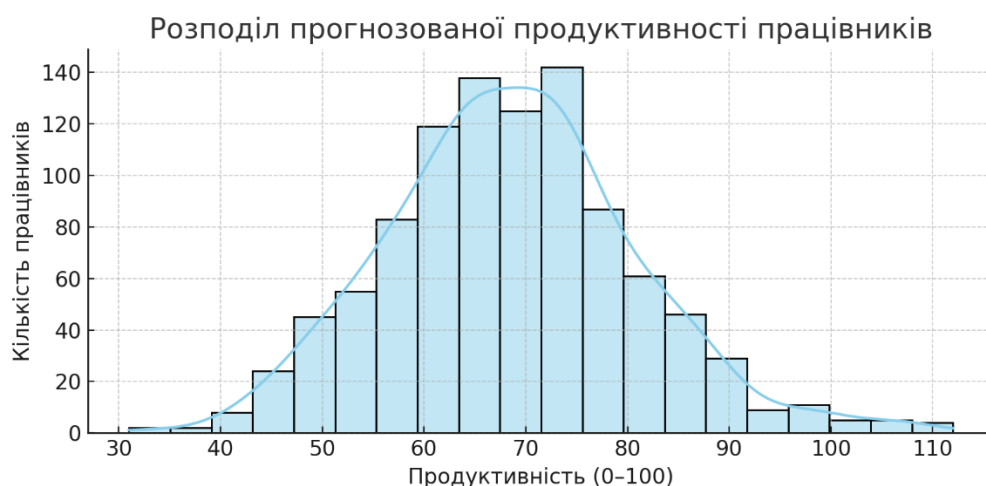


Рисунок 5.6 – Розподіл прогнозованої продуктивності працівників (самостійне виконання)

Однак, не менш важливим є наявність відхилень від норми – зокрема, працівників, чий рівень продуктивності опускається нижче 50 або піднімається вище 90. У реальних умовах це може вказувати:

- на ознаки вигорання, демотивації або проблеми з балансом роботи й життя (низькі значення);
- або навпаки – на високу мотивацію, залученість і готовність до професійного зростання (високі значення).

Інтерпретація таких крайових значень має особливе значення для HR-практик, зокрема при:

- розподілі навантаження в команді;
- ухваленні рішень про бонуси, перегляд KPI або преміювання;
- виявленні потенційних лідерів і «ризикових» працівників.

Окрім цього, візуалізація може бути використана як інструмент комунікації з керівництвом. Наприклад, графік легко адаптується для щотижневих або щомісячних звітів про стан команди, що дозволяє оцінити динаміку продуктивності з часом, виявити ефект змін (нових правил, навантаження, переходу на гібридну модель роботи тощо).

Таким чином, включення цього графіку до функціоналу системи – це не лише технічна особливість, а й потужний аналітичний інструмент, який дозволяє не

просто бачити дані, а діяти на їх основі. Це повністю відповідає сучасній концепції data-driven HR, де управлінські рішення мають базуватися на об'єктивних аналітичних висновках, а не інтуїції.

Крім цього, інтерфейс системи формує графік порівняння поточного працівника із середніми показниками по компанії. Цей візуальний компонент доступний безпосередньо на сторінці результатів аналізу, під текстовими рекомендаціями (див. рис. 5.7). Такий підхід дозволяє зробити висновки на основі контексту, а не лише абсолютних значень.



Рисунок 5.7 – Порівняння працівника із середніми значеннями по компанії (самостійне виконання)

На графіку відображено чотири ключові параметри:

- продуктивність;
- добробут;
- вік;
- місячний дохід.

Кожен параметр представлено у двох стовпчиках – блакитний (поточний працівник) та жовтий (середнє значення серед працівників компанії). Це дає змогу миттєво оцінити відхилення. Наприклад, якщо добробут або продуктивність

помітно нижчі за середні, це є сигналом до втручання, навіть якщо формально значення ще не критичні.

Такий графік дозволяє не лише візуально порівняти працівника із «середнім портретом», але й підвищує обґрунтованість прийняття рішень. Це особливо важливо в умовах великих колективів, де персональний аналіз кожного працівника є ресурсозатратним.

Таким чином, обидві візуалізації – як розподіл продуктивності, так і порівняльний аналіз – є невід’ємною частиною інтерфейсу, що допомагає не лише зробити прогноз, а й надати аналітичне обґрунтування для управлінських рішень.

ВИСНОВКИ

У даній магістерській роботі було досліджено підходи до аналізу продуктивності працівників, що працюють дистанційно, а також запропоновано систему підтримки HR-рішень, яка поєднує методи машинного навчання для оцінки ризику звільнення, прогнозування продуктивності та генерації персоналізованих рекомендацій.

На основі аналізу сучасного стану дистанційної роботи в IT-галузі встановлено, що зростаюча гнучкість у графіках і форматах зайнятості створює нові виклики для управління персоналом. Зокрема, питання виявлення прихованого вигорання, зниження мотивації або високої плинності кадрів набули особливої актуальності.

Для досягнення мети дослідження було:

- зібрано та попередньо оброблено два відкриті набори даних;
- сформовано цільові змінні для двох задач: прогнозування продуктивності та ризику звільнення;
- побудовано три моделі машинного навчання (регресійну, класифікаційну, кластеризаційну);
- реалізовано веб-застосунок, що інтегрує зазначені моделі в єдину систему.

Результати експериментального тестування показали достатню якість моделей ($F1\text{-score} = 0.85$ для класифікації, $R^2 = 0.76$ для регресії), що дозволяє використовувати їх у практиці HR-аналітики. Демонстраційні сценарії засвідчили, що система здатна:

- виявляти потенційно проблемних працівників до настання критичної ситуації;
- диференціювати поведінкові кластери персоналу;
- надавати інтерпретовані та контекстно релевантні рекомендації.

Таким чином, у ході роботи було доведено, що використання моделей машинного навчання може значно підвищити ефективність ухвалення рішень у сфері управління персоналом у дистанційному форматі. Розроблена система має перспективу подальшого розвитку – шляхом масштабування на більші обсяги

даних, включення нових параметрів (наприклад, командної динаміки) або розширення функціоналу для менеджерів.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Buffer. (2023). State of Remote Work. URL: <https://buffer.com/state-of-remote-work/2023> (дата звернення: 01.04.2025).
2. Owl Labs. State of Remote Work 2022. URL: <https://owllabs.com/state-of-remote-work> (дата звернення: 01.04.2025).
3. Benefits of Working From Home (Working Remotely). URL: <https://www.flexjobs.com/blog/post/benefits-of-remote-work> (дата звернення: 07.04.2025).
4. IBM Security. Cost of a Data Breach Report 2023. URL: <https://www.ibm.com/reports/data-breach> (дата звернення: 09.04.2025).
5. The rise in remote work since the pandemic and its impact on productivity. URL: https://www.bls.gov/opub/btn/volume-13/remote-work-productivity.htm?utm_source=chatgpt.com (дата звернення: 19.04.2025).
6. Study finds hybrid work benefits companies and employees. URL: https://news.stanford.edu/stories/2024/06/hybrid-work-is-a-win-win-win-for-companies-workers?utm_source=chatgpt.com (дата звернення: 24.04.2025).
7. Are remote workers more productive? That's the wrong question. URL: https://stackoverflow.blog/2023/11/27/are-remote-workers-more-productive-that-s-the-wrong-question/?utm_source=chatgpt.com (дата звернення: 26.04.2025).
8. Remote workers' well-being: Are innovative organizations really concerned? A bibliometrics analysis. URL: https://www.sciencedirect.com/science/article/pii/S2444569X24001343?utm_source=chatgpt.com (дата звернення: 28.04.2025).
9. Tech Neck Is Back. URL: https://www.wired.com/story/tech-neck-wfh-health/?utm_source=chatgpt.com (дата звернення: 05.05.2025).
10. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Декларативно-темпоральний підхід до побудови пояснень в інтелектуальних інформаційних системах. Вісник Нац. техн. ун-ту "ХПІ": зб. наук. пр. Темат. вип. Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПІ». 2020. № 2(4). С. 51-56. DOI: 10.20998/2079-0023.2020.02.09 (дата звернення: 07.05.2025).

11. Chalyi, Sergii & Leshchynskyi, Volodymyr. (2023). Інформаційна технологія оцінки пояснень в інтелектуальній інформаційній системі. Системи управління, навігації та зв'язку. Збірник наукових праць. 4. 120-124. DOI: 10.26906/SUNZ.2023.4.120 (дата звернення: 07.05.2025).

12. Dataset «Remote Work Productivity». URL: <https://www.kaggle.com/code/alaabdelstar/remote-work-productivity> (дата звернення: 15.05.2025).

13. Dataset «IBM HR Analytics Employee Attrition & Performance». URL: <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset> (дата звернення: 15.05.2025).

14. GitHub [Електронний ресурс]: Архів кваліфікаційної роботи магістра. - Режим доступу: https://github.com/YuliiaMaster/2025_M_IPZm-23-4_Archive/ (дата звернення: 12.06.2025).

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ

10. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Декларативно-темпоральний підхід до побудови пояснень в інтелектуальних інформаційних системах. Вісник Нац. техн. ун-ту "ХПІ": зб. наук. пр. Темат. вип. Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПІ». 2020. № 2(4). С. 51-56. DOI: 10.20998/2079-0023.2020.02.09 (дата звернення: 07.05.2025).

11. Chalyi, Sergii & Leshchynskyi, Volodymyr. (2023). Інформаційна технологія оцінки пояснень в інтелектуальній інформаційній системі. Системи управління, навігації та зв'язку. Збірник наукових праць. 4. 120-124. DOI: 10.26906/SUNZ.2023.4.120 (дата звернення: 07.05.2025).