

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
Харківський національний університет радіоелектроніки

Центр післядипломної освіти  
Кафедра Програмної інженерії

**КВАЛІФІКАЦІЙНА РОБОТА**  
**Пояснювальна записка**

другий (магістерський)  
(рівень вищої освіти)

Дослідження методів автоматизованого аналізу резюме в  
рекрутингових системах

Виконала:

студентка групи ІПЗзdm-22-1  
Таран А. О.  
(прізвище, ініціали)

Спеціальність 121 – Інженерія  
програмного забезпечення

Тип програми Освітньо-наукова  
Керівник доц., к.т.н, Кириченко І. В.  
(посада, прізвище, ініціали)

Допускається до  
захисту  
Зав. Кафедри

\_\_\_\_\_

З. В. Дудар

2024 р.

Харківський національний університет радіоелектроніки

Факультет Центр післядипломної освіти

Кафедра Програмної інженерії

Рівень вищої освіти другий (магістерський)

Спеціальність 121 – Інженерія програмного забезпечення

(код і повна назва спеціальності)

Тип програми освітньо-наукова програма

Освітня програма Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ р.

## ЗАВДАННЯ

### НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту групи ІПЗзДМ-22-1 Таран Анастасії Олександрівні

1. Тема роботи: “Дослідження методів автоматизованого аналізу резюме в рекрутингових системах”

затверджена наказом по університету від 22.04.2024 р. No 60Стз

2. Термін подання студентом роботи до екзаменаційної комісії 10.06.2024р.

2. Вихідні дані до роботи: дослідити методи автоматизованого аналізу резюме за допомогою методів машинного навчання.

4. Перелік питань, що потрібно опрацювати в роботі: мета роботи, аналіз предметної галузі та постановка задачі, огляд та аналіз літературних джерел з дослідження, методи вирішення проблеми, дослідження теоретичне, дослідження практичне, аналіз отриманих результатів, опис прототипу додатку, висновки.

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів	Примітка
1	Аналіз предметної галузі	01.04.2024 - 14.04.2024	Виконано
2	Визначення методів автоматизованого аналізу резюме	08.04.2024 - 15.04.2024	Виконано
3	Підготовка до експерименту та збір даних	16.04.2024 - 20.04.2024	Виконано
4	Проектування архітектури та розробка прототипу	21.04.2024 - 25.04.2024	Виконано
5	Проведення експерименту	26.03.2024 - 27.04.2024	Виконано
6	Підготовка пояснювальної записки	28.04.2024 - 03.06.2024	Виконано
7	Підготовка презентації та доповіді	04.06.2024 - 05.06.2024	Виконано
8	Перевірка на плагіат	11.06.2024	Виконано
9	Перевірка на нормоконтроль	16.06.2024	Виконано
10	Оцінка роботи рецензентом	17.06.2024	Виконано
11	Попередній захист роботи	19.06.2024	Виконано
12	Здача роботи у електронний архів	20.06.2024	Виконано
13	Допуск до захисту завідуючим кафедри	20.06.2024	Виконано

Дата видачі завдання «01» квітня 2024 р.

Студент \_\_\_\_\_ Таран А.О.

Керівник роботи \_\_\_\_\_ доц. каф. ПІ Кириченко І.В.

## РЕФЕРАТ/ABSTRACT

Кваліфікаційна робота магістра містить: 91 с., 35 рис, 9 табл., 53 джерела.

АНАЛІЗ РЕЗЮМЕ, ВІДБІР КАНДИДАТІВ, МЕТОДИ МАШИННОГО НАВЧАННЯ, НАЙМ КАНДИДАТІВ, РЕКРУТИНГ, СИСТЕМИ ВІДСТЕЖЕННЯ КАНДИДАТІВ, ШТУЧНИЙ ІНТЕЛЕКТ, NAIVE BAYES, RANDOM FOREST, SVM.

Об'єкт дослідження - процеси автоматизованого аналізу резюме в рекрутингових системах.

Предмет дослідження - аналіз методів машинного навчання для вирішення задач пов'язаних з автоматизацією аналізу резюме в рекрутингових системах.

Мета дослідження - оптимізація роботи спеціалістів з найму шляхом автоматизації фільтрації резюме.

Методами розробки та проектування є аналіз проблемної області дослідження, вибір методу автоматизованого аналізу резюме шляхом вирішення задачі багатокритеріального вибору та проведення експерименту.

У результаті роботи було запропоновано та реалізовано три моделі навчання для автоматизованого аналізу резюме, по одній для кожного з обраних методів машинного навчання: Naive Bayes, SVM, Random Forest. Розроблено прототип програмної системи, який дозволяє порівнювати результати їхньої роботи.

RESUME ANALYSIS, SELECTION OF CANDIDATES, MACHINE LEARNING METHODS, HIRING CANDIDATES, RECRUITING, APPLICANT TRACKING SYSTEMS, ARTIFICIAL INTELLIGENCE, NAIVE BAYES, RANDOM FOREST, SVM.

The object of the study is the processes of automated analysis of resumes in recruiting systems.

The subject of research is artificial intelligence methods for automating resume analysis in recruiting systems.

The purpose of the study is to optimize the work of specialists with the highest way of automating resume filtering.

The methods of development and design are the analysis of the problem area of research, the choice of the method of automated analysis of resumes based on the solution of the problem of multi-criteria selection conducting an experiment.

As a result of the work, 3 models were created, one for each of the machine learning methods: Naive Bayes, SVM, Random Forest, and the results of their work were compared.

Умови публікації пояснювальної записки

Я, Таран Анастасія Олександрівна

(прізвище, ім'я, по батькові)

студентка групи ІІЗЗдм-22-1 здобувач вищої освіти на другому (магістерському) рівні

кафедра програмної інженерії

(повна назва кафедри)

заявляю: моя кваліфікаційна робота на тему «Дослідження методів автоматизованого аналізу резюме», що буде представлена до ЕК для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу ElArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайоmlена 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.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	8
ВСТУП.....	9
1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ.....	12
1.1 Аналіз предметної області.....	12
1.2 Огляд існуючих ATS та їхньої функціональності.....	15
1.3 Дослідження цінності використання штучного інтелекту у рекрутингових системах.....	31
Висновки до Розділу 1.....	37
2 МЕТОДИ ВИРІШЕННЯ ЗАДАЧІ.....	39
2.1 Вибір методу обробки резюме.....	39
2.2 Вирішення задачі багатокритеріального вибору.....	45
Висновки до Розділу 2.....	49
3 ПРОВЕДЕННЯ ЕКСПЕРИМЕНТУ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ.....	50
3.1 Визначення вхідних даних та опис алгоритму.....	50
3.2 Аналіз отриманих результатів.....	56
Висновок до Розділу 3.....	59
ВИСНОВКИ.....	60
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	62
ДОДАТОК А.....	68
ДОДАТОК Б.....	69
ДОДАТОК В.....	70
ДОДАТОК Г.....	73
ДОДАТОК Д.....	89
ДОДАТОК Е.....	90

## **ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ**

ATS - Applicant Tracking System

AI - Artificial intelligence

HRM(S) - Human Resources Management(System)

ML - Machine Learning

NLP - Natural language processing

ІІ - Штучний інтелект

## ВСТУП

Після настання пандемії COVID-19 світ IT-рекрутингу почав стрімко змінюватися здебільшого на користь кандидатів: віддалена робота, стрімкий розвиток IT напрямку (навіть незважаючи на те, що багато сфер бізнесу, які виступали замовниками програмного забезпечення, сильно постраждали), попит на IT спеціалістів стрімко збільшувався і ринок все більше переходив на сторону кандидата. Велику кількість відгуків від кандидатів можна було очікувати лише на позиції Intern/Trainee рівня, адже стали популярними різноманітні IT курси і “світчинг” в IT став звичною справою. Все це сприяло стрімкій популяризації роботи в IT, та зумовило тенденцію, що досвідчених кандидатів сильно не вистачало.

З початком повномасштабного вторгнення ситуація стрімко змінилася. Через небезпеку для життя розробників, можливість їхньої мобілізації, стрімке збільшення кібер атак та ризику втрати чутливої інформації, багато замовників почали виходити з українського ринку та шукати альтернативні варіанти та ринки праці. Частими умовами для компаній стали можливість найму або українців за межами країни, або локалів в інших країнах. Міжнародні компанії, які раніше конкурували в Україні, стали боротися за першість і відкривати офіси в країнах Європи та Латинської Америки.

Через це ринок праці в Україні стрімко обвалився, велика кількість спеціалістів залишилася без роботи. У результаті на ту, порівняно невелику кількість вакансій, що залишилася, посипалися сотні, а в деяких випадках, навіть тисячі резюме кандидатів. На превеликий жаль, це вплинуло не лише на початкові позиції, а на всі рівні і направленості.

У зв'язку з цим у рекрутерів сильно зросло навантаження на первинному етапі, а саме відборі резюме. Багато кандидатів, у пошуках хоч якогось варіанту, не завжди читають опис вакансії або не зважають на нього в надії, що наймаюча сторона все ж зацікавиться кандидатурою апліканта та запросить на співбесіду. У зв'язку з цим рекрутери витрачають вагому частину свого робочого часу, щоб із купи заявок відібрати ті, що все ж відповідають заявленим вимогам [1].

Підсумовуючи вищесказане, можна зробити висновок, що тема дослідження методів оптимізації рекрутингових процесів є *актуальною*, особливо для ІТ бізнесів, адже це дозволить більш раціонально та доцільно розподіляти час та навантаження рекрутерів, а також сприятиме удосконаленню процесу найму в компанію. Оскільки пошук та найм нових працівників є не тільки часозатратним, але й потребує залучення чималих коштів, запропоновані в результаті проведеного дослідження методи автоматизованого аналізу резюме можуть бути використані для покращення вже існуючих систем відстежування кандидатів або створення нових систем. Використання цих методів також опосередковано впливатиме і на розподілення та доцільне використання фінансових ресурсів компаній.

Враховуючи вищеописану проблематику, *метою* дослідження буде аналіз методів автоматизованої фільтрації резюме, що спрямовано на оптимізацію роботи спеціалістів з найму. Це в свою чергу, дозволить зменшити витрати часу під час відбору резюме та помилки при роботі з відгуками за допомогою методів інтелектуального аналізу даних.

Для виконання поставленої мети необхідним є виконання наступних *задач*:

- провести огляд та аналіз літературних джерел та існуючих систем відстеження кандидатів;
- дослідити цінність використання машинного навчання у рекрутингових системах;
- визначити та порівняти найпопулярніші методи, які можуть використовуватися для автоматизованого аналізу резюме;
- визначити метрики, що будуть використані для оцінки ефективності застосування обраних методів;
- вибрати оптимальний метод для автоматизованого аналізу резюме за допомогою вирішення задачі багатокритеріального вибору та експериментального дослідження.

Об'єкт дослідження - процеси автоматизованого аналізу резюме в рекрутингових системах.

Предмет дослідження - аналіз методів машинного навчання для вирішення задач пов'язаних з автоматизацією аналізу резюме в рекрутингових системах.

Під час проведення дослідження було використано аналітичні та емпіричні методи аналізу даних. Шляхом аналізу літературних джерел [2, 3, 4, 5, 6] та сучасного стану дослідження подібної тематики, було обрано 3 методи машинного навчання та проведена їхня оцінка за допомогою вирішення задачі багатокритеріального вибору. Для підтвердження отриманих результатів було реалізовано три моделі для автоматизованого аналізу резюме, по одній для кожного з обраних методів машинного навчання та проведено експерименти на обраному датасеті.

У результаті дослідження було проаналізовано роботу методів машинного навчання та обрано найоптимальніший. Цей метод можна застосовувати у сучасних ATS для вдосконалення їхньої роботи та автоматизації відбору резюме на етапі відгуку на вакансію.

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

### 1.1 Аналіз предметної області

Під час дослідження було проаналізовано біля 60 наукових публікацій та профільних інтернет джерел [7, 8, 9], що так чи інакше пов'язані з підбором та наймом персоналу, семантичним аналізом тексту, а також автоматизацією процесу найму за допомогою методів машинного навчання.

Процес відбору кандидатів на позицію відрізняється від компанії до компанії та залежить від багатьох факторів: рівень розвитку та розмір фірми, наявність побудованих процесів, цінності та місія компанії, тощо. Проте, можна говорити про класичний процес відбору працівників (рис. 1.1), який поділяється на 6 етапів та в тому чи іншому вигляді присутній у кожній організації [10]:

- виявлення потреби у співробітниках;
- розробка вимог до кандидата, його портрету;
- пошук кандидатів за допомогою використання різних джерел;
- збір інформації про кандидатів на посаду: аналіз резюме, первинне інтерв'ю з рекрутером, анкетування тощо;
- оцінювання кандидатів з використанням традиційних та нетрадиційних методів та процедур: інтерв'ю, тестове завдання, поліграф, тощо;
- прийняття рішення про найм.

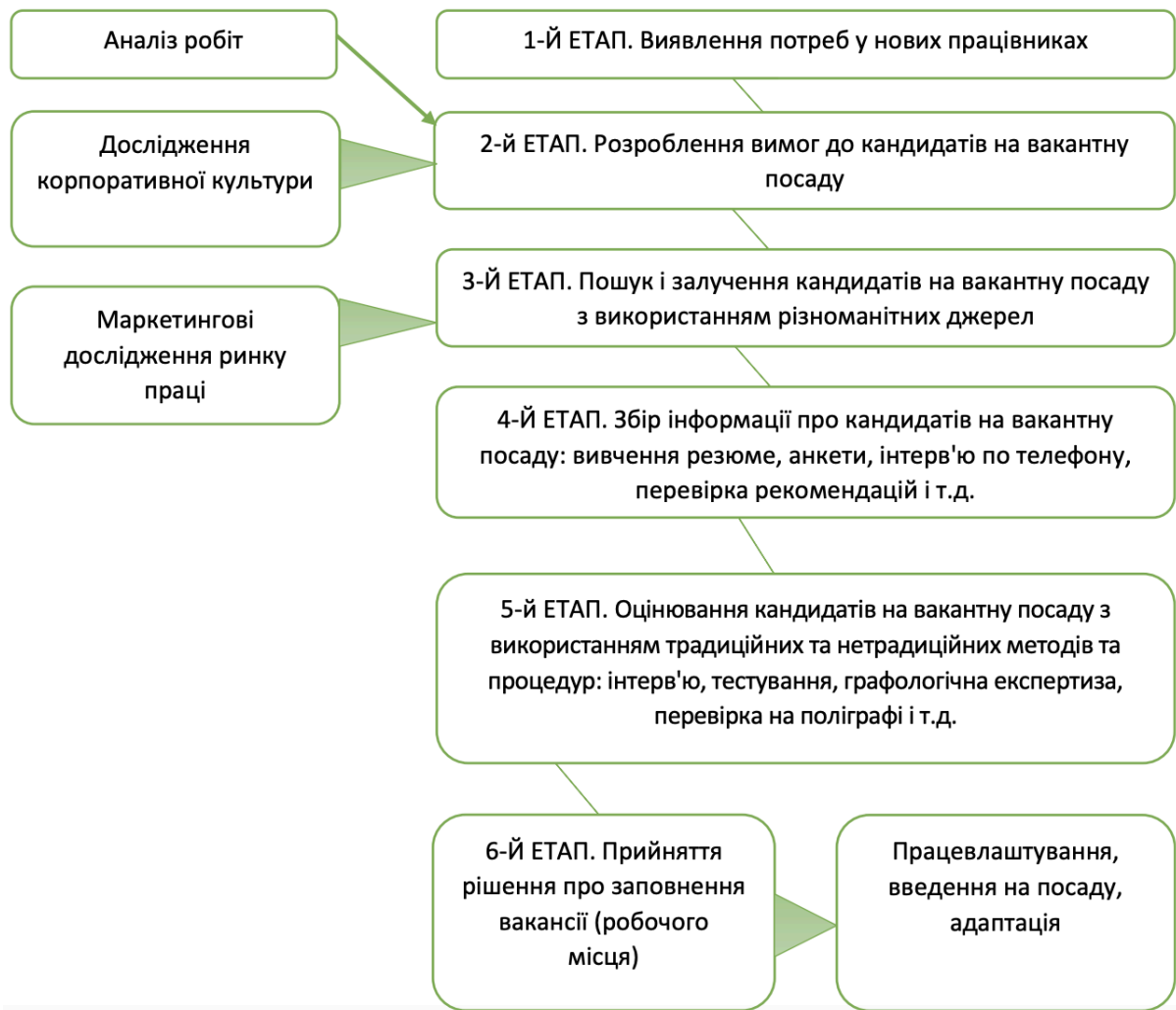


Рисунок 1.1 - Класичний процес відбору працівників [10]

Під час дослідження наукових джерел [11, 12, 13] було помічено закономірність, що найбільша кількість з розглянутих робіт присвячена лише огляду існуючих Applicant Tracking System (ATS). В деяких статтях, додатково проводилися порівняння характеристик ATS [14, 15].

ATS (система відстеження кандидатів) - це програмне забезпечення, яке допомагає організаціям у вирішенні рекрутингових завдань. ATS може бути як додатком, так і онлайн-сервісом, в залежності від потреб і розміру організації. Існують також безкоштовні ATS та ATS з відкритим програмним кодом [16].

Також слід відмітити, що в багатьох наукових публікаціях [17, 18] окремо відзначена проблема упередженості (bias) та різноманітності (diversity) під час процесу найму на роботу. Що, в свою чергу, є основою для вивчення питання

щодо використання машинного навчання, як засобу який допомагає проводити неупереджену оцінку кандидатів, а також дозволяє зменшити відмови кандидату через власні погляди та переконання рекрутера щодо гендеру, раси, віку, національності, сімейного чи соціального статусу тощо.

В деяких дослідженнях, відзначається, що використання різних класичних методів автоматизованого аналізу резюме, навпаки, може породжувати неявну, а інколи й приховану дискримінацію при прийнятті рішень. Наприклад, в статті “Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision making in the context of HR recruitment and HR development” [17] розглядається припущення, що дискримінація шляхом алгоритмічного прийняття рішень здебільшого не вивчена в контексті управління людськими ресурсами, та алгоритмічне прийняття рішень може призвести до несправедливого ставлення до певних груп людей, неявної дискримінації та несправедливості.

Варто зазначити, що навчені моделі ШІ дійсно може роботи помилки і деякі резюме можуть бути помилково пропущені, або навпаки, не надані на розгляд через те, що тема ще недостатньо вивчена, але здебільшого, це може бути через недоліки алгоритмів, а не через упереджену оцінку кандидатів. Все ж таки необхідно зазначити, що системи ШІ, по своїй природі не схильні до мізантропії, на відміну від людей.

Частина наукових статей присвячена способам аналізу резюме та їхньому розподілу по вакансіям в залежності від тексту в резюме [19, 20]. Наприклад, семантичний аналіз тексту є одним з найбільш успішних та ефективних напрямів, для машинного пошуку інформації, визначення сенсу текстових документів, пошуку залежностей, тощо [19]. Також пропонуються варіанти формування анотацій та коротких вижимок через парсинг резюме. Наприклад, у статті “Intelligent Hiring with Resume Parser and Ranking Using Machine Learning and Natural Language Processing” автори пропонують власну систему, яка ранжує резюме будь-якого формату відповідно до заданих обмежень або вимог, наданих

фірмою-клієнтом, використовуючи обробку природної мови (NLP) і машинне навчання (ML) [21].

Не менш цікавим варіантом автоматизації рекрутингових процесів є збір резюме з мережі Інтернет та холодна розсилка кандидатам з використанням чат-ботів. Також, деякі автори відзначають можливість використання ШІ під час співбесіди та аналіз голосу та обличчя співрозмовника [22]. Аналіз цих досліджень значно розширює сприйняття можливостей автоматизації рекрутингових процесів, проте, вже не входить в рамки цього дослідження, так як під час дослідження буде розглянуто лише етап аналізу та відбору резюме рекрутером.

## 1.2 Огляд існуючих ATS та їхньої функціональності

Процес найму складається з кількох етапів, які організації виконують для залучення, оцінки та відбору нових працівників. Він зазвичай включає публікацію вакансій, відбір кандидатів, перевірку резюме, проведення початкових співбесід, тестування та перевірку рекомендацій, остаточні співбесіди, надання пропозицій роботи та їх прийняття. Хоча цей процес може варіюватися в залежності від розміру компанії, її культури та типу посади, основна мета залишається незмінною: знайти найкращого кандидата і забезпечити його успішний онбординг в компанію [23].

Для автоматизації робочого процесу та допомоги при відборі кандидатів на ринку існує доволі багато різних засобів. Компанії пропонують ATS, які допомагають зручно відстежувати наявних кандидатів, ATS з AI, що допомагають виконувати фільтрацію кандидатів та пришвидшують пошук, роблять автоматичні розсилки з пропозицією вакансій і навіть проводять первинний скринінг. Є Human Resources Management Systems (HRMS), в яких вбудовані і опції для рекрутингу.

Проаналізувавши декілька рейтингів ATS систем [24, 25], було зроблено висновок, що єдиної думки, щодо їхньої популярності не існує. Так, у таблиці 1.1.

зазначено рейтинги найкращих ATS у 2023 за даними Forbes Advisor [24] та SelectSoftware REVIEWS [25].

Таблиця 1.1 - Найкращі ATS у 2023 за даними Forbes Advisor та SelectSoftware REVIEWS (Таблиця виконана самостійно)

<i>№ n.n</i>	<i>Forbes Advisor</i>	<i>SelectSoftware REVIEWS</i>
1	Freshteam by Freshworks	Rippling
2	JazzHR	Pinpoint
3	BreezyHR	BambooHR
4	Rippling	Recruitee
5	Greenhouse	ClearCompany
6	Zoho Recruit	Teamtaylor
7	BambooHR	Manatal
8	Workable	Ashby
9	Bullhorn	Greenhouse
10	Recruit CRM	VidCruiter
11	Recruitee	Lever
12	-	Trakstar Hire
13	-	Avature
14	-	JazzHR
15	-	SmartRecruiters
16	-	iCIMS

Всі вказані в таблиці системи пропонують приблизно однаковий функціонал роботи з профілями кандидатів. Та деякі з них мають розширений функціонал взаємодії з резюме. Далі буде розглянуто декілька ATS, що використовують штучний інтелект.

BreezyHR [26] - ця система дозволяє не тільки створювати та публікувати вакансії, сортувати кандидатів в залежності від етапу на якому вони знаходяться,

працювати зі скоркардами, але й має певний функціонал для автоматичної взаємодії з кандидатами. Breezy має так званий “Stage Action”, тобто допомагає автоматизувати процес в залежності від етапу, на якому знаходиться кандидат: від надсилання повідомлень, анкет і оцінок до збору відгуків команди та планування співбесід.

Breezy дає можливість проводити попередній пре-скрін кандидатів без участі в ньому рекрутера за допомогою анкетування. Система автоматично просуває на наступний етап, або дискваліфікує кандидатів, в залежності від їхніх відповідей на питання анкети. Таким чином, рекрутер може приділити більше часу потенційно найкращим кандидатам і не витратити час на тих, хто завідомо не підходить. Також, є можливість налаштування “Match Score”. За допомогою цієї функції можна налаштовувати параметри для кожної вакансії і система буде оцінювати кожен відгук, що надсилається на вакансію. Таким чином наймаюча команда відразу бачить, які з резюме найбільше відповідають вимогам вакансії та зможе не витратити час на резюме, які отримали найнижчий рейтинг, що в разі скорочує роботу з відгуками. Проте ця функція оплачується додатково і є доволі дорогою для українського ринку - \$165 на місяць. При цьому вона, є не дуже зручною, адже налаштовується у вигляді шкали: який аспект має яку значимість [26]. На рисунках 1.2 - 1.4 наведено графічний інтерфейс BreezyHR.

**Candidate Scoring** [Preview](#)

Scorecards define custom criteria your Hiring Team will use to evaluate candidates. [Learn More](#)

**Available Scorecards**

Choose which Team Scorecard to use with this position.

Scorecard (optional)

[+ Add Scorecard](#)

**Match Score**

Configure the category weighting to use when evaluating candidates for their Match Score.

Enabled Disabled

Least Important Most Important

Education

Work History

Skills

Industries

Languages

Certifications

Executive

Management

[Save Changes](#)

Рисунок 1.2 - Match Score [26]

**breezy**

**Matt Watkins** ★

3.7 Match Score 7.5 Team Score

---

**Matt Watkins** ▼

Full Stack Javascript Developer @ ...

1mo ago 2m ago

**Senior Software Engineer**

United States

Interviewing  Applied

Smart Match	Sort	Matt Watkins	Darren Bounds
Match Score		3.7	1.3
Skills		2 <sup>7/10</sup>	1 <sup>7/10</sup>
Work History		3 <sup>7/10</sup>	0 <sup>7/10</sup>
Education		0 <sup>7/10</sup>	0 <sup>7/10</sup>
Industries		9 <sup>7/10</sup>	7 <sup>7/10</sup>
Management		10 <sup>7/10</sup>	0 <sup>7/10</sup>
Team Scorecard	Sort		
Team Score		7.5	7.5

« Previous Candidates Page 1 of 1 Next Candidates »

All Candidates (Admin) Columns 134 candidates More Candidate Email [+ Add Candidate](#)

Photo	Name	Match	Score	Email Address	Phone Number	Address	Desired Salary	Position / Pool
	Matt Watkins	3.7	7.5	[Redacted]	[Redacted]	[Redacted]	[Redacted]	Senior Software Engineer
	Kevin Bragers	3.7	5.5	[Redacted]	[Redacted]	[Redacted]	[Redacted]	Dongle Delegate

Рисунок 1.3 - Приклади результатів оцінювання [26]

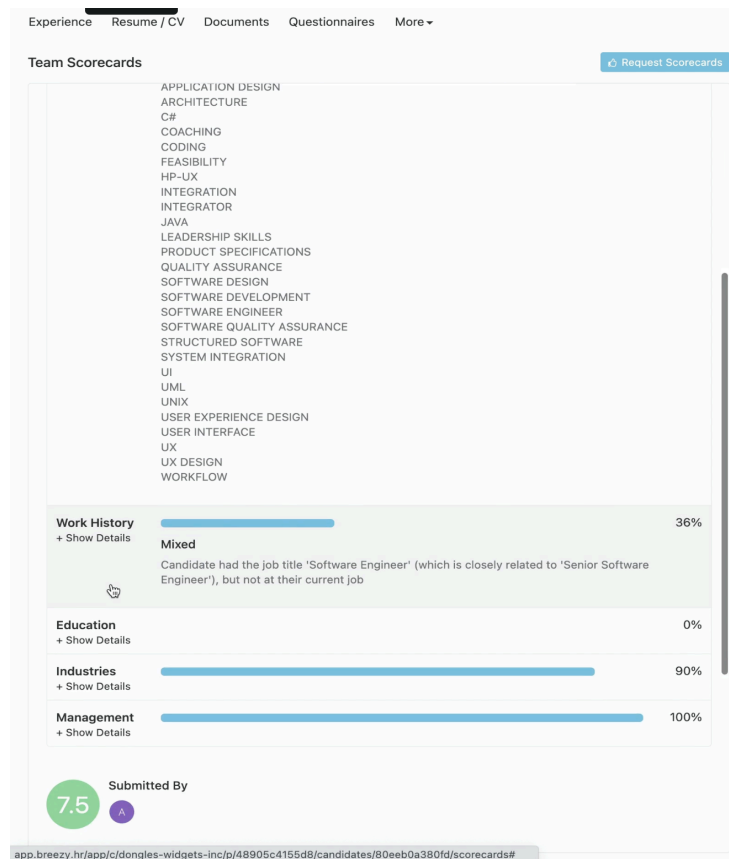


Рисунок 1.4 - Розподілення даних [26]

Схожу до Match Score функцію має Manatal [27].

Manatal - це рекрутингове програмне забезпечення з використанням штучного інтелекту, створене для ефективного пошуку та найму кандидатів. Воно спрощує процес найму, надаючи найкращих кандидатів для певних посад [27].

У Manatal можна задати параметри і система буде сортувати кандидатів відповідно до вимог та виділяти найбільш підходящих. Інтерфейс та реалізація теж подібні до Breezy та наведені на рисунку 1.5 [27].

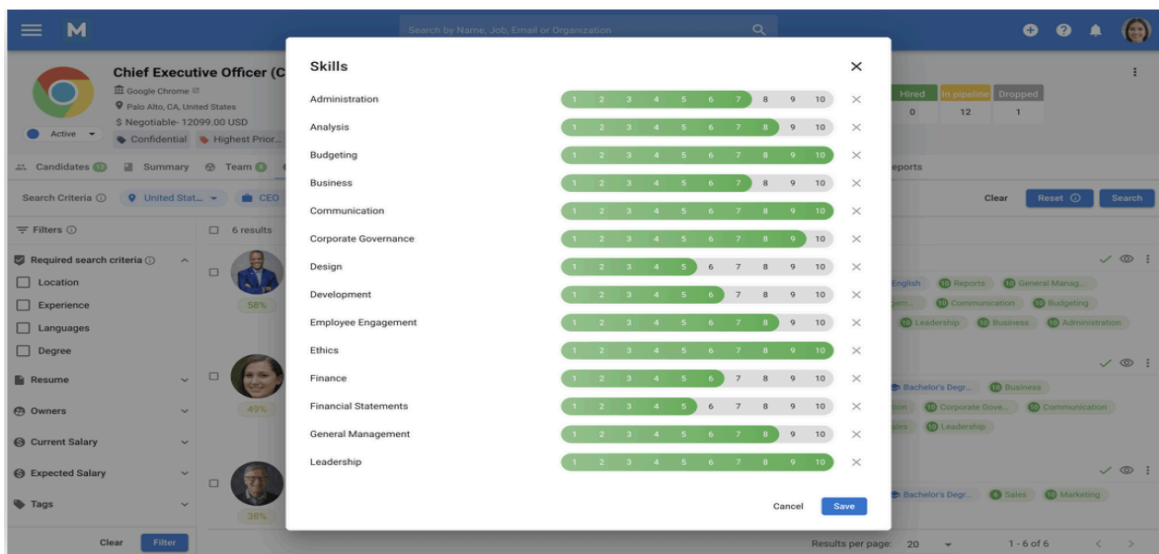


Рисунок 1.5 - Оцінка важливості навичок [27]

GreenHouse надає своїм користувачам систему, яка, за їхнім офіційним промо, забезпечує чесний та справедливий процес найму. Ця система допомагає боротися з несвідомими упередженнями на кожному етапі найму за допомогою підказок, інструментів вимірювання та інших засобів. Використовуючи GreenHouse, можна дозволити кандидатам записувати правильну вимову своїх імен (рис. 1.6) і вказувати особисті займенники, щоб рекрутингова команда могла правильно їх використовувати, забезпечуючи кандидатам відчуття поміченості та прийняття [28].

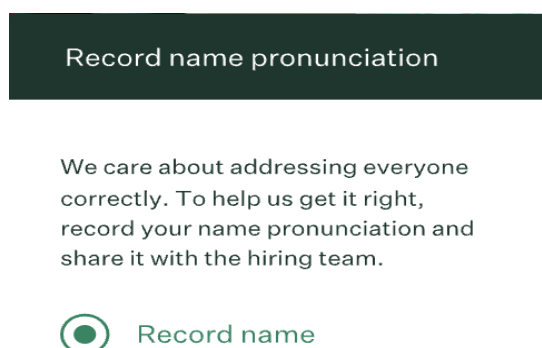


Рисунок 1.6 - Запис правильної вимови імені [28]

GreenHouse також має функцію з автовідмовою: рекрутер до кожної вакансії може додати невеличке опитування (рис. 1.7) і призначити автовідмову на певний варіант відповіді. При цьому рекрутер має можливість переглянути всі заявки, яким система надіслала відмову автоматично [28].

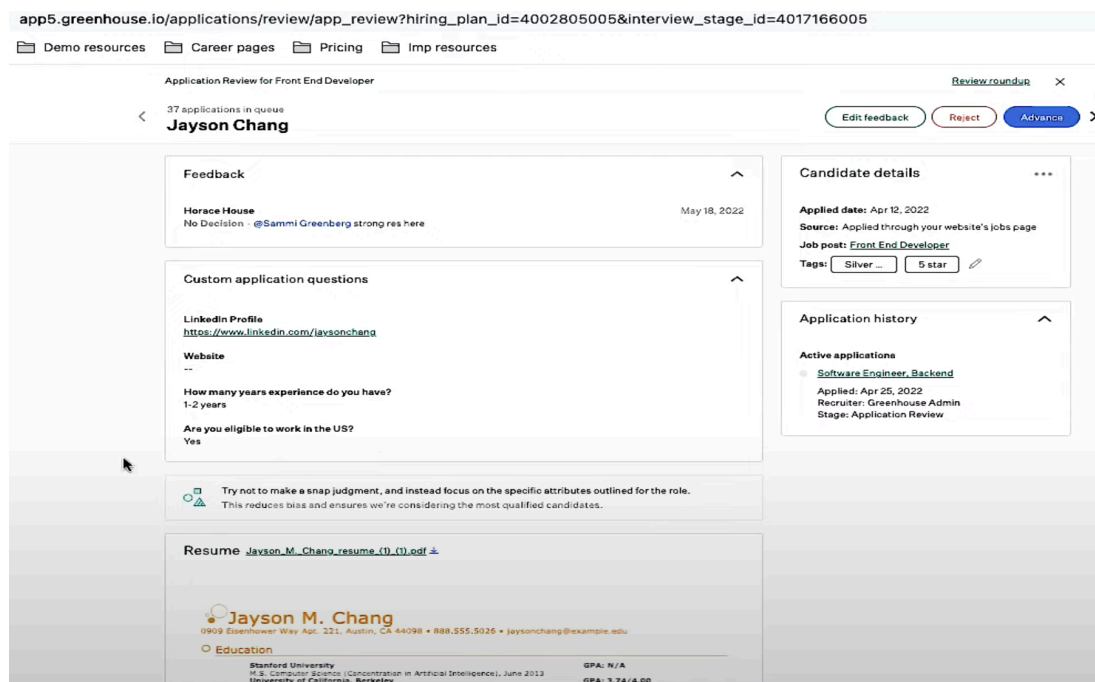


Рисунок 1.7 - Створення опитування (GreenHouse) [28]

Схожу фічу з анкетуванням має і Avature [29].

Avature ATS надає повний набір інструментів для надання інноваційних, цифрових і повністю індивідуальних послуг з підбору персоналу для підприємства [29].

З метою зменшити навантаження на рекрутера та заощадити час менеджерів, у Avature пропонують свої інструменти перевірки, які оцінюють та ранжують кандидатів на основі попередньо визначених критеріїв. Спеціаліст з найму може налаштувати запитання для будь-якої вакансії, щоб швидко визначити кваліфікованих кандидатів і дозволити системі подбати про автоматичний розподіл на інші вакансії тих, хто не потрапив у відбір. Слід зазначити, що у даній системі відсутня функціональність надсилання авто-відмови, вона лише ранжує кандидатів на основі їхніх відповідей [29].

SmartRecruiters надає своїм клієнтам ефективну систему відстеження кандидатів (ATS) з інтегрованими маркетинговими інструментами для найму, які підвищують продуктивність рекрутерів, покращують командну співпрацю та забезпечують позитивний досвід кандидатів [30].

Крім того, SmartRecruiters пропонує продукт під назвою SmartAssistant, який представляє собою рекрутингове програмне забезпечення зі штучним інтелектом. Це ПЗ автоматично відбирає найкращих кандидатів із наявної бази, аналізує кандидатів з використанням статистики та оцінює їх за необхідними навичками (рис. 1.8). Завдяки функції парсингу резюме, це ПЗ забезпечує ефективний спосіб виявлення ключових слів, навичок та інших важливих критеріїв, що допомагає відфільтрувати велику кількість заявок та визначити найкращих кандидатів [30].

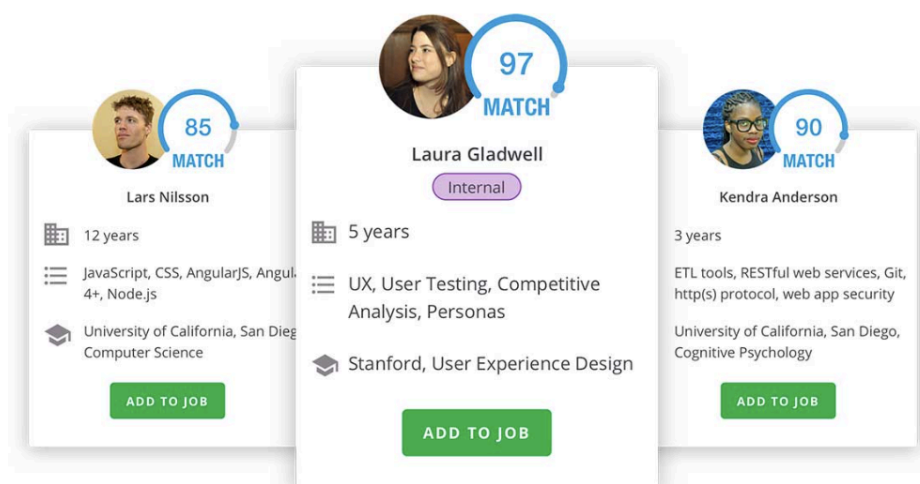


Рисунок 1.8 - Відбір та оцінка кандидатів (SmartAssistant) [30]

iCIMS пропонує набирати та утримувати таланти за допомогою своєї платформи найму на основі штучного інтелекту, розробленої для підвищення ефективності, скорочення витрат на підбір персоналу та перевищення очікувань кандидатів [31].

На сайті компанії запевняють, що їхня ATS з використанням AI допоможе зосередитися на найперспективніших кандидатах, адже надається можливість керувати кількістю кандидатів та розставляти пріоритети за допомогою

автоматичного порівняння кандидатів, ранжування та співставлення з вимогами вакансії [31].

Zoho Recruit [32] пропонує потужні ATS і CRM на єдиній рекрутинговій платформі.

Для організацій, які прагнуть оптимізувати свій процес найму Zoho Recruit пропонують рішення на основі ШІ. Завдяки програмі Zia's Candidate Matching можна швидше підібрати потрібного кандидата на відкриту посаду на основі його досвіду та навичок і порівняти його з іншими кандидатами (рис 1.9). Функція парсингу резюме (рис. 1.10), що працює на основі ШІ та власних алгоритмів, допомагає стандартизувати та узгодити структуру резюме відповідно до форми, яка прийнята в компанії. Також є функція з оцінкою навичок та надання балів кандидатам, рекрутеру треба лише заповнити форму перед публікацією вакансії та вказати, які навички є "Must have", а які "Nice to have" (рис. 1.11). Крім цього, Zoho Recruit має чат бота, який сповіщає кандидатів про їхній статус та інформує про наявні вакансії (рис. 1.12) [32].

The screenshot displays the Zoho Recruit interface for a job opening titled "Marketing Executive" at "Pinnacle Corp". The interface is divided into several sections:

- Navigation Menu:** Home, My Actions, Job Openings, Candidates, Interviews, Clients, Contacts, Campaigns, Reports, Dashboards, Forecasts.
- Job Opening Summary:** Shows 0 Screening, 0 Submissions, 0 Interview, and 0 Offered.
- Job Opening Details:**
  - Job Opening Status: In-progress
  - Modified Time: a while ago
  - Contact Name: Eduardo Vargas
  - Target Date: Jul 10
  - City: San Antonio
  - Assigned Recruiter(s): Aaron Brown
  - Territory: Austin
- Job Opening Information:**
  - Posting Title: Marketing Executive
  - Contact Name: Eduardo Vargas
  - Assigned Recruiter(s): Aaron Brown
  - Target Date: Jul 10
  - Job Opening Status: In-progress
  - Industry: Marketing
  - Zip/Postal Code: 3601
  - Country: United States
  - Salary: 20000
  - Created By: Patricia Boyle
- Job Opening ID Details:**
  - Job Opening ID: ZR\_L\_JOB
  - Client Name: Pinnacle Corp
  - Account Manager: Patricia Boyle
  - Date Opened: Apr 3
  - Job Type: Full time
  - City: San Antonio
  - State/Province: Washington
  - Work Experience: 1-3 years
  - Skillset: Marketing
  - Modified By: Patricia Boyle
- Forecast Details:**
  - Number of Positions: 10
  - Expected Revenue: \$ 0.00
  - Revenue per Position: \$ 0.00
  - Actual Revenue: \$ 0.00
- Candidate Matches (Zia):**
  - 3 Matches (Associate All)
  - Jack Wilson: 87% Match, Marketing Executive, 2 Years Experience
  - Quinn Rivers: 78% Match, quinn.rivers@gmail.com
  - Wendy Ortiz: 74% Match, B.Tech, Fresher

Рисунок 1.9 - Порівняння резюме кандидатів (Zoho Recruit) [32]

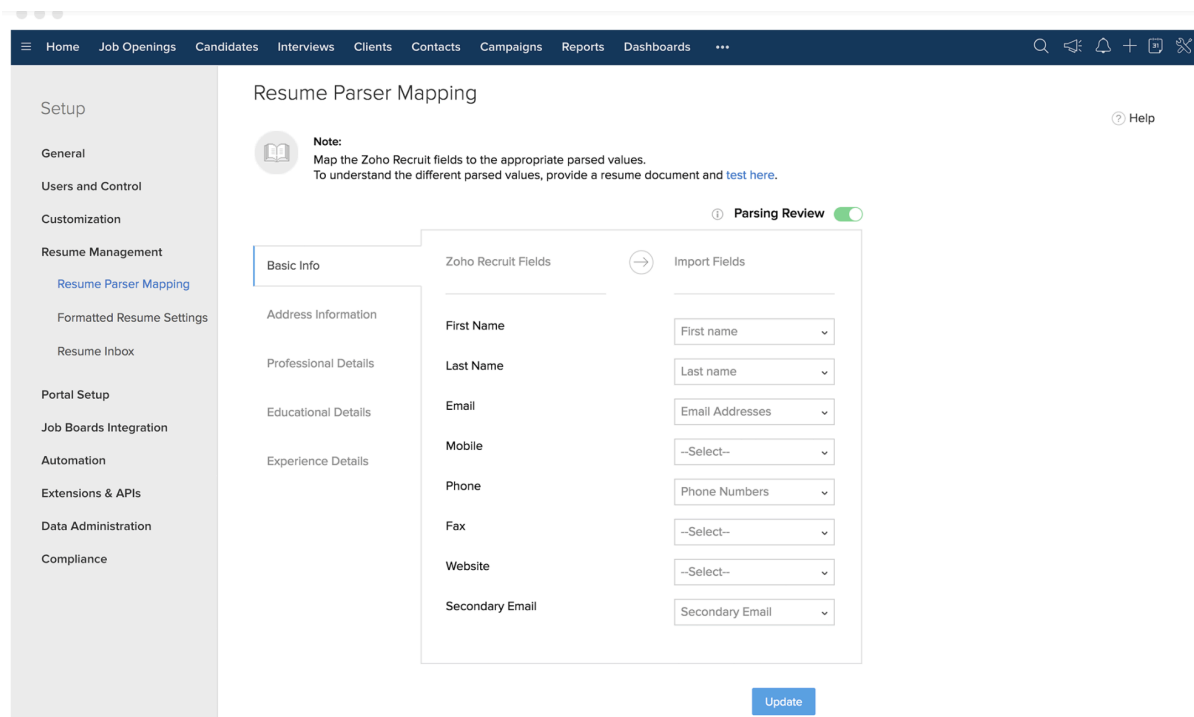


Рисунок 1.10 - Парсинг резюме (Zoho Recruit) [32]

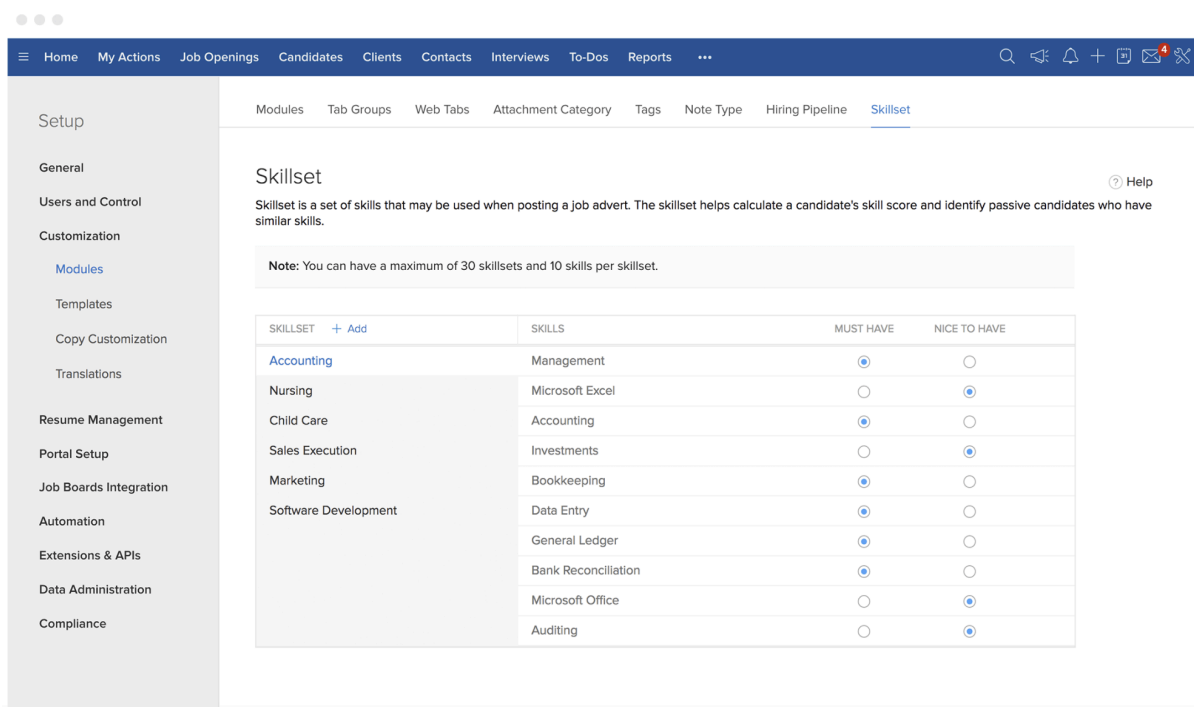


Рисунок 1.11 - Оцінка навичок кандидатів (Zoho Recruit) [32]

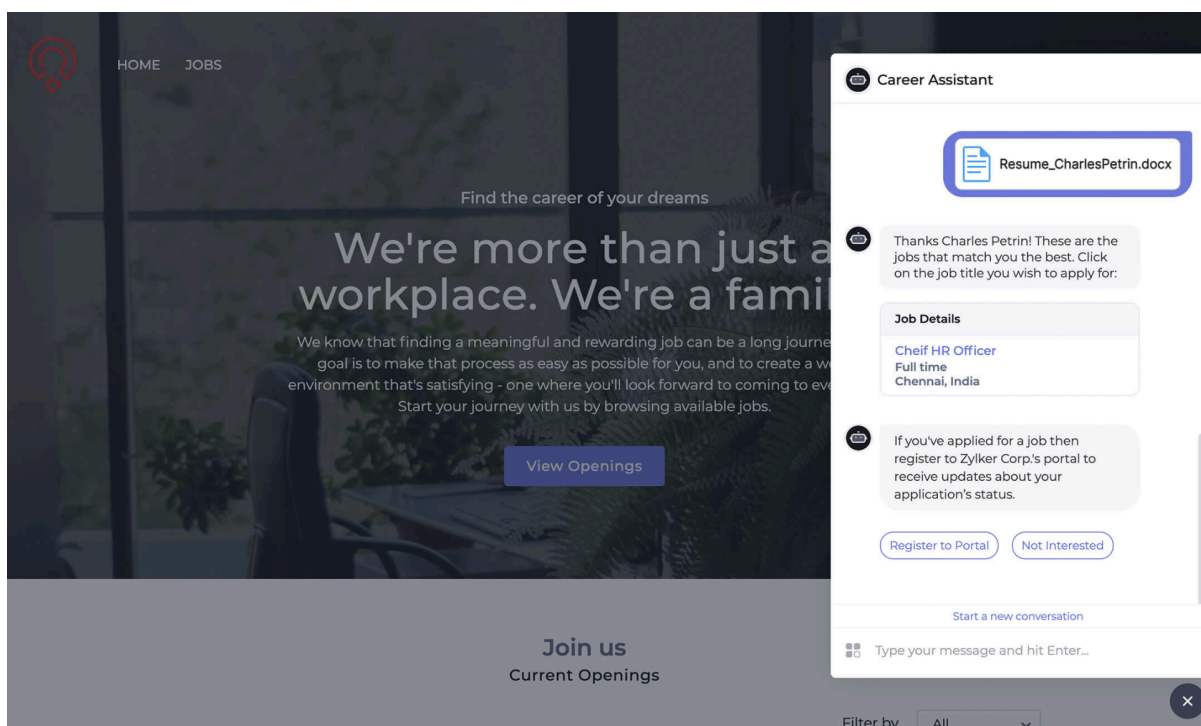


Рисунок 1.12 - Чат бот (Zoho Recruit) [32]

Слід відмітити, що не всі з вище перерахованих систем популярні в українському IT. Так, згідно з опитуванням рекрутерів, опублікованому на сайті Happy Monday [33], учасники опитування у своїй роботі найчастіше використовують такі системи управління кандидатами:

- CleverStaff — 19%.
- Excel — 14,6%.
- ATS власного виробництва — 12,7%.
- Breezy — 6,1%.
- Hurma — 6,1%.
- Greenhouse — 5,4%.
- Zoho — 5,4%.

Як можна побачити, деякі рекрутери на ринку українського IT й досі не користуються трекінговими системами для автоматизації своєї роботи. Найчастіше ці рекрутери користуються в своїй роботі звичайними Excel таблицями, ще частина респондентів використовують в своїх компаніях ATS власного виробництва. Крім цього, можемо побачити 2 системи, яких немає в

рейтингах на міжнародних ресурсах, але якими користуються рекрутери на ринку українського ІТ – це CleverStaff [34] та Hurma [35], які є, власне, українськими продуктами.

CleverStaff [34] вважається найпопулярнішим софтом для автоматизації рекрутингу в Україні, що підтверджується опитуванням, що наведено вище. Система має гейміфікацію для рекрутерів, що спрощує процес навчання користування системою, а також є додатковим мотиваційним механізмом, адже в процесі використання рекрутер отримує бали за досягнення у використанні ATS. CleverStaff має функцію парсингу резюме, а також систему автопідбору за допомогою ІІ, яка присвоює бали кандидатам та сортує їх в залежності від того, наскільки резюме апліканта відповідає вимогам вакансії (рис. 1.13).

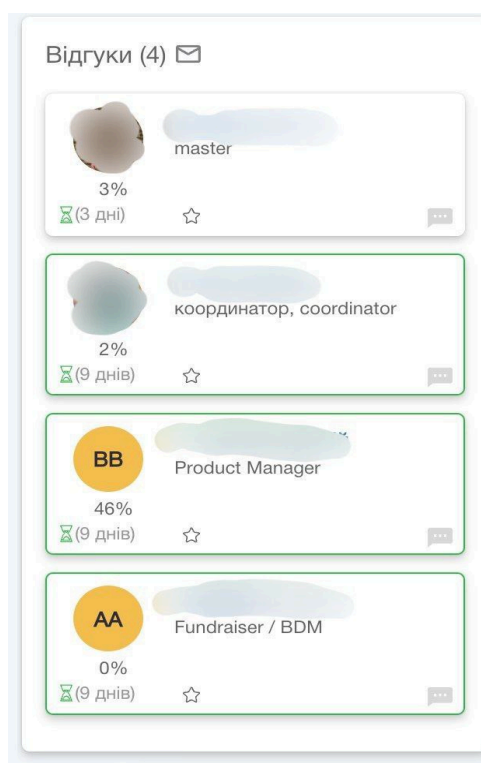


Рисунок 1.13 - Присвоєння балів кандидатам в CleverStaff (Рисунок виконано самостійно)

Відповідно до власного досвіду роботи з CleverStaff можна відзначити, що ця система ранжування діє доволі неякісно і часто можна зіткнутися з тим, що у нерелевантних кандидатів може бути високий відсоток співпадіння, а у

підходящих, навпаки, низький. Система продовжує сортувати кандидатів, навіть, коли параметри не вказані, що скоріше заважає роботі рекрутера (рис. 1.14).

Вид зайнятості 🔄 повна зайнятість	Кількість позицій 👤 (не вказано)	Сфера діяльності 🏢 (не вказано)	Створена 📅 22 Травня Таран А.
Локація 📍 Україна, Київ	Кінцевий термін 📅 (не вказано)	Оціночна картка ★ <a href="#">Картка за замовчуванням</a>	Бюджет 💰 (не вказано)
Зарплата 💵 (не вказано)	Шаблон відгука 📄 Шаблон за замовчуванням	Оплата 💳 (не вказано)	

<b>Вимоги до кандидата</b>			
Обов'язкові навички (не вказано)	Буде плюсом (не вказано)	Мови (не вказано)	Досвід роботи 3 роки

Рисунок 1.14 - Незаповнений профіль вакансії в CleverStaff (Рисунок виконаний самостійно)

Hurma [35] також є української ATS та HRM платформою. Проте, ШІ вони в своїй системі поки що не мають. Тобто, це звичайний інструмент автоматизації, який дозволяє публікувати вакансії, якісно вести базу кандидатів за допомогою канбан-дошки (рис. 1.15), надсилати кандидатам листи, при переводі їх на наступний етап, та робити звітність по роботі над вакансіями (рис. 1.16).

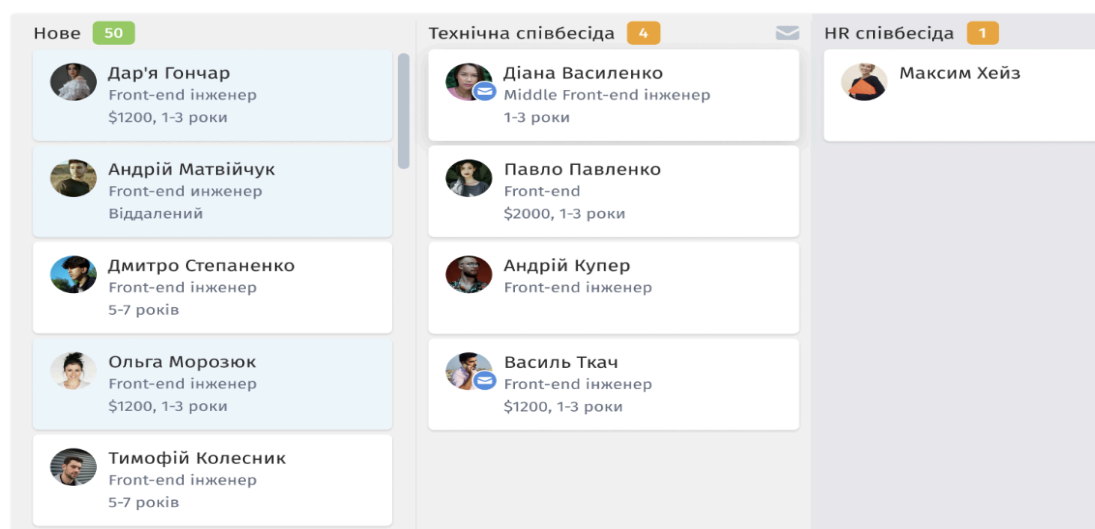


Рисунок 1.15 - Інтерфейс ATS Hurma [35]

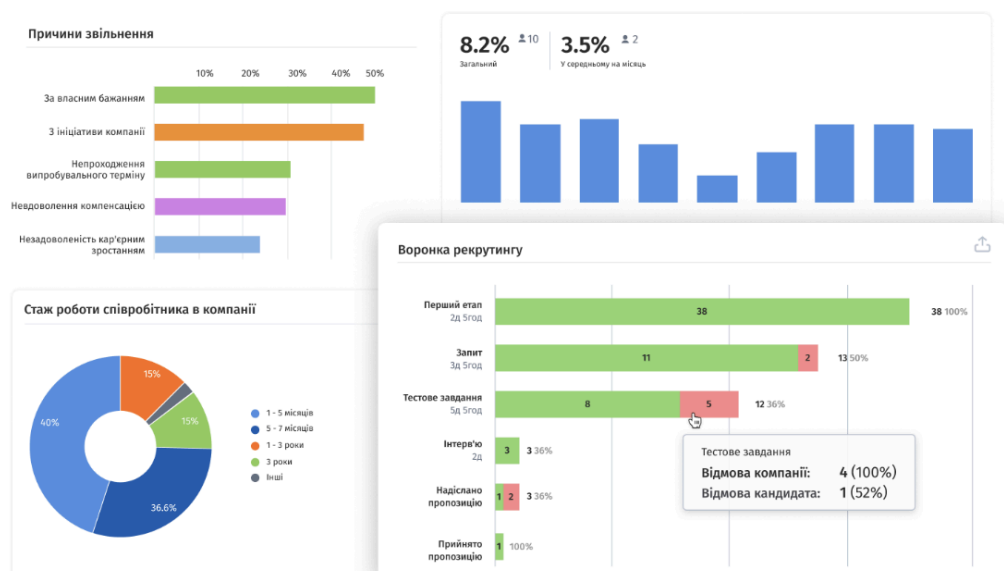


Рисунок 1.16 - Функція звітності (Hurma) [35]

Провівши власне невелике опитування у соціальній мережі LinkedIn, до вказаного списку хотілось би ще додати Workable [36], як популярну систему серед багатьох компаній на українському ринку.

Workable [36] це іноземне ПЗ як послуга (SaaS) надає систему відстеження кандидатів та інструменти для рекрутингу, що допомагають керувати та підтримувати процес найму. Workable використовує ШІ, але це стосується опису вакансії (рис 1.17) та пошуку пасивних кандидатів (рис. 1.18), функціонал, що обробляє відгуки у складі системи зараз відсутній.

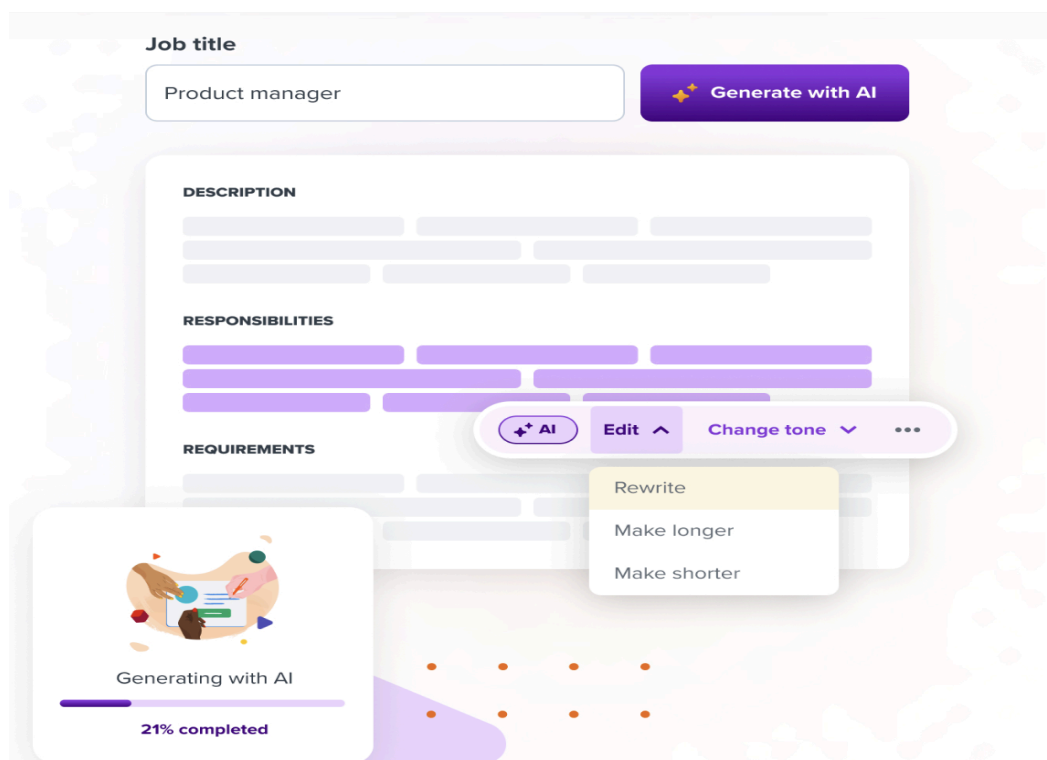


Рисунок 1.17 - Створення опису вакансії за допомогою ШІ (Workable) [36]

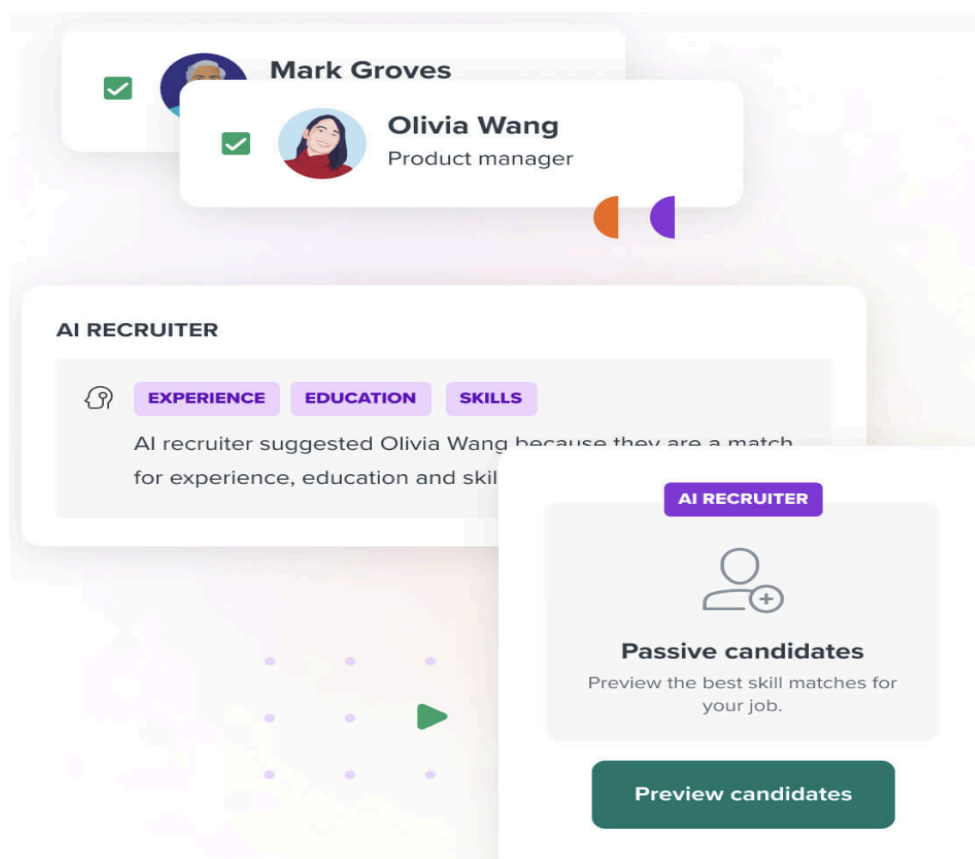


Рисунок 1.18 - Пошук пасивних кандидатів за допомогою ШІ (Workable) [36]

Результати порівняння ATS, якими найчастіше користуються спеціалісти з рекрутингу в Україні, наведено в таблиці 1.2.

Таблиця 1.2 - Порівняльний аналіз ATS (Таблиця виконана самостійно з використанням даних з [12] та [15])

	CleverStaff	Breezy	Hurma	GreenHouse	Zoho	Workable
Інтерфейс	веб-браузер	веб-браузер, iOS, Android	веб-браузер	веб-браузер	веб-браузер, iOS, Android	веб-браузер, iOS, Android
Наявність пробної версії	+	+	+	за запитом	+	+
Постинг вакансій	+	+	+	+	+	+
Перенесення резюме з інших ресурсів	+	+	+	+/-	+	+/-
Пошук кандидата в базі компанії за визначеними критеріями	+	+	+	+	+	+
Ранжування кандидатів за обраними критеріями	+	+	-	-	+	-
Авториджект кандидатам, якщо вони не відповідають ключовим критеріям	-	+	-	+	-	-
Наявність чат-бота	-	-	-	-	+	-
Запрошення кандидата на співбесіду	+	+	+	+	+	+
Планування співбесіди	+	+	+	+	+	+
Тестування психологічних рис кандидата та його оцінка	-	-	+/-	-	-	-

Продовження таблиці 1.2

	CleverStaff	Breezy	Hurma	GreenHouse	Zoho	Workable
Отримання зворотного зв'язку від менеджерів та керівництва по кандидатам	+	+	+	+	+	+
Повідомлення про відмову або прийом на роботу	+	+	+	+	+	+
Аналіз виконаної праці, звіти для керівництва	+	+	+	+	+	+

Відповідно до проведеного аналізу, можна зробити висновок, що ATS системи, досить популярні на ринку праці в Україні та допомагають автоматизувати процес, але не завжди задовольняють всі головні потреби рекрутера при відборі кандидатів на певну вакансію. Також необхідно відмітити, те, що не багато з них використовують штучний інтелект для обробки резюме і не всі спрямовані на пришвидшення роботи рекрутера, через складний інтерфейс та специфічні налаштування (інколи час на встановлення налаштувань більший за ефективність роботи системи).

### 1.3 Дослідження цінності використання штучного інтелекту у рекрутингових системах

За результатами аналізу літературних джерел та існуючих систем відстеження кандидатів можна зробити висновок, що тема автоматизації рекрутингу та використання штучного інтелекту в процесі найму є актуальною в наш час. Науковці та компанії-розробники ATS пропонують різноманітні методи автоматизації процесів та, не тільки пришвидшення роботи з кандидатами на різних етапах, але і підвищення якості цієї роботи. Далі, на основі проведеного аналізу, розглянемо ціннісну пропозицію дослідження.

Перш за все, необхідно зазначити, що дослідження розраховане виключно на споживачів серед ІТ індустрії, оскільки найм кандидатів керується законодавством та має свою специфіку у кожній із сфер. Тому в подальшому ці напрацювання можуть використовуватися для удосконалення ATS та використання спеціалістами з найму персоналу, агентства та ІТ компанії, що наймають ІТ спеціалістів.

Ціннісна пропозиція - це сукупність переваг, які організація пропонує своїм споживачам та зацікавленим сторонам. Деякі ціннісні пропозиції можуть бути інноваційними та пропонувати нові або революційні рішення. Інші можуть бути схожими на вже існуючі на ринку, але з додатковими функціями та особливостями. [37].

Далі розглянемо більш детально, що саме можна запропонувати можливим користувачам.

Цінності, що пропонуються:

- автоматизація процесу відбору резюме на етапі відгуку на вакансію;
- оптимізація рекрутингового процесу та часу спеціалістів з найму за рахунок автоматизації;
- зменшення вірогідності прийняття невірної рішення;
- зменшення вартості кожного найму.

Розглянемо ці цінності, чи відповідають вони реальності та чи є важливими.

Автоматизація процесу відбору резюме на етапі відгуку дозволяє рекрутеру сфокусуватися на подальших етапах відбору та виборі найкращого кандидата. Це стало більш актуально за повномасштабного вторгнення, адже значно зріс рівень безробіття та стрімко збільшилася кількість відгуків.

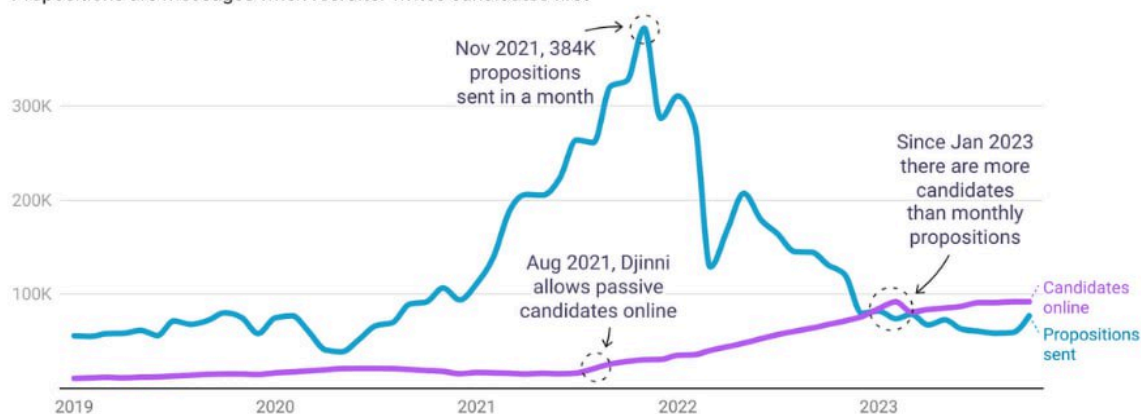
В підтвердження цьому свідчать статистичні дані, що пропонує популярний український ресурс для пошуку роботи – Djinni [38].

Так, за їхньою інформацією кількість кандидатів у пошуку на Djinni з 2019 виросла у 9 разів. В свою чергу кількість пропозицій кандидатам від рекрутерів зараз повернулася до рівня 2019-го після того, як, з настанням COVID-19, відбувся величезний ріст у 2021-му [38].

Наприкінці 2021 в середньому один кандидат міг отримувати 16 пропозицій на місяць (при тому, що кандидатів, які шукали роботу було близько 30 000). Зараз людей, що знаходяться в пошуку нової роботи - 92000 і рекрутери пропонують роботу приблизно третій частині з цих кандидатів. Середня кількість повідомлень на кандидата онлайн зараз менше 1 (рис. 1.19) [38].

### Candidates online grow x9 since 2019, recruiters' propositions fall to 2019 levels post-2021 rise

Propositions are messages when recruiter writes candidates first



### Propositions per candidate online

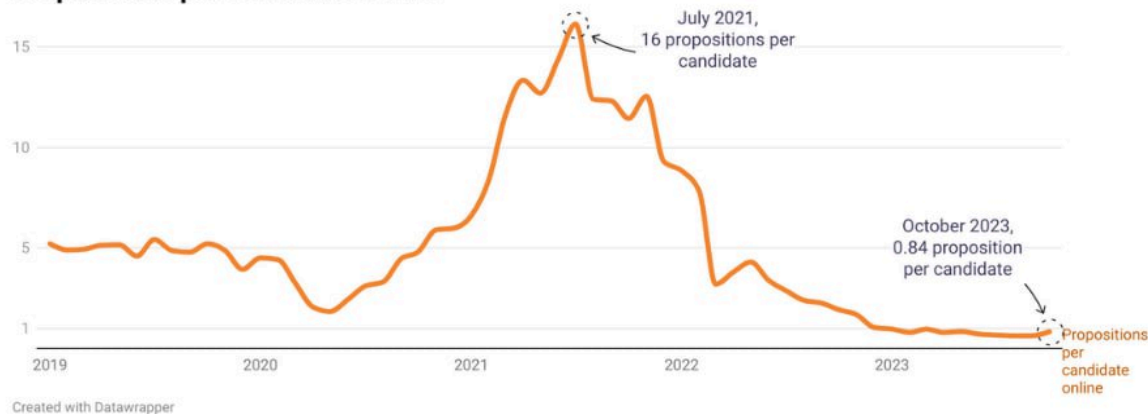


Рисунок 1.19 - Порівняння попиту-пропозиція за 2019-2023 рр. [38]

Станом на кінець грудня 2023 року кількість вакансій онлайн становить 6,8 тисяч, що на 30% менше порівняно з кінцем 2022 року. Найбільше падіння спостерігалось в серпні, як і в інших показниках. Число кандидатів у пошуку на Djinni досягло 92,6 тисяч, що на 20% більше порівняно з кінцем минулого року, але лише на 300 більше, ніж три місяці тому. Минулого року щоквартально

додавалося по 10 тисяч кандидатів. Конкуренція в усіх категоріях зростає з 7,9 до 13,5 протягом року (рис. 1.20) [38].

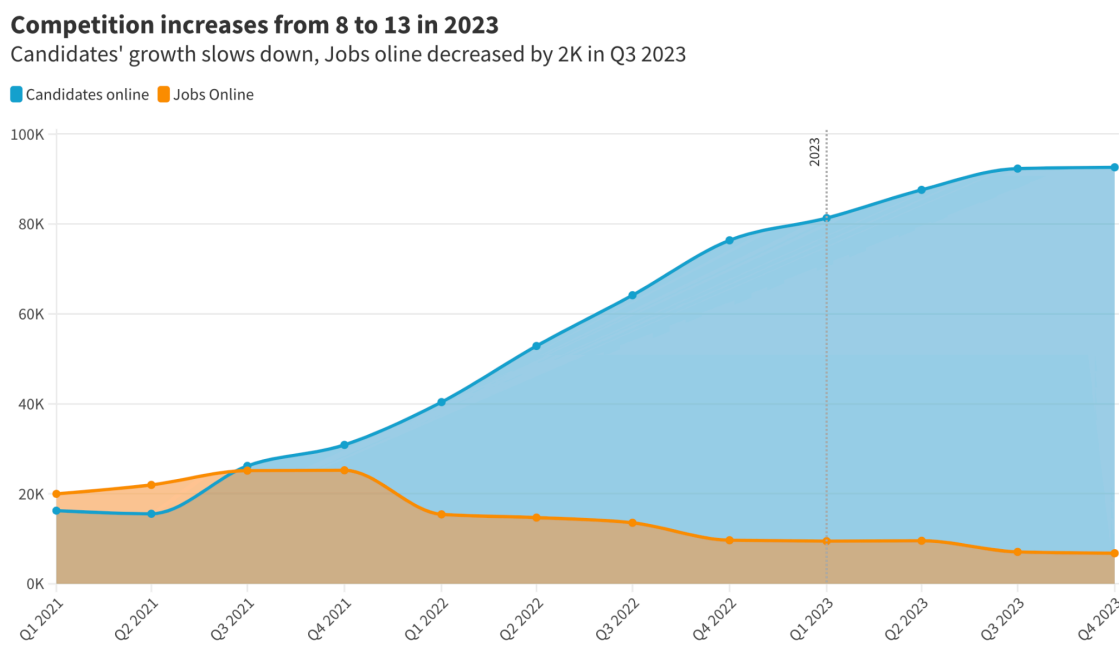


Рисунок 1.20 - Співвідношення опублікованих вакансій до кандидатів, 2023р. [38]

Згідно з цими даними, можна побачити, що попит значно перевищує пропозицію, що стає додатковим навантаженням для рекрутерів при обробці всіх відгуків.

Та чому саме первинний відбір резюме потребує оптимізації та не розглядається оптимізація інших етапів як більш важлива?

Справа в тому, що на цьому етапі рекрутерам зустрічається найбільше нерелевантних кандидатів та обробка усіх заявок може займати декілька днів.

В підтвердження цього буде наведено приклади з власного досвіду, що зображені на рисунках 1.21 та 1.22, де показані вже закрита вакансія, та вакансія в процесі роботи відповідно.

## IT HR People Partner Accepted Offer

26 Aug, 2022 · DP · Brazil

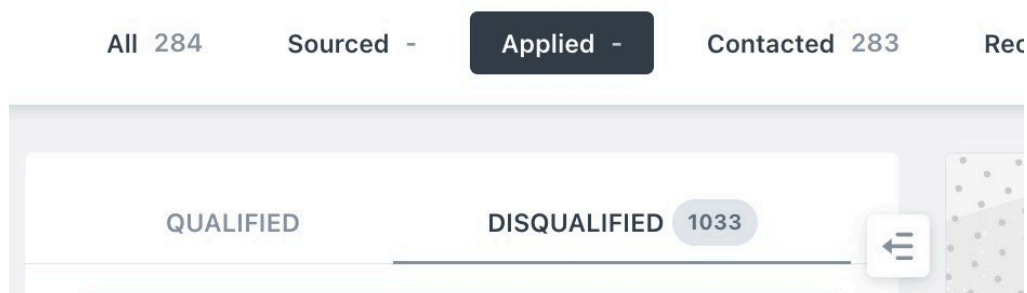


Рисунок 1.21 - Кількість відмов після перевірки резюме, закрита вакансія (Рисунок виконаний самостійно)

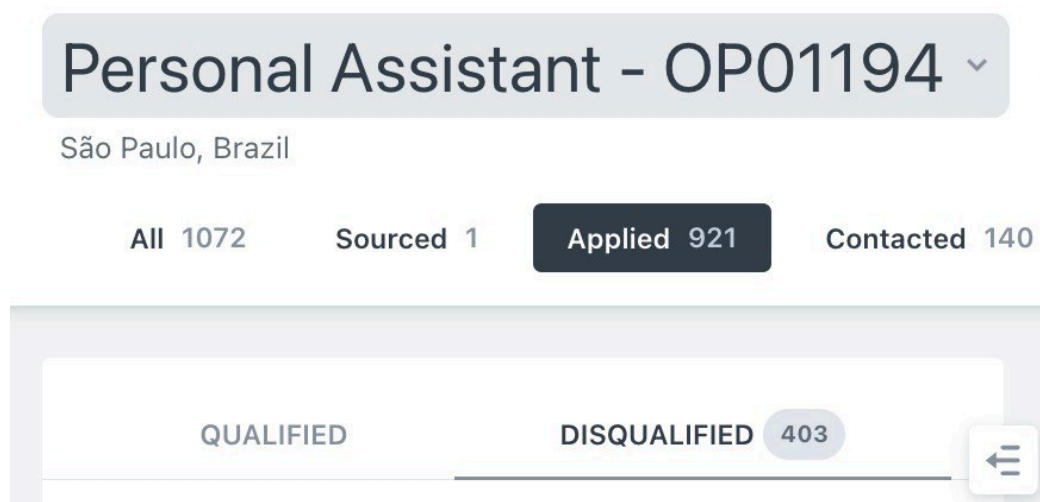


Рисунок 1.22 - Кількість відмов після перевірки резюме, відкрита вакансія (Рисунок виконаний самостійно)

Як можна бачити з прикладу на рисунку 2.3. під час обробки відгуків, рекрутер відмовив більше ніж 1 тисячі кандидатів, що відгукнулися на вакансію. На рисунку 2.4. зазначено, що відмова надіслана 403 кандидатам, але ще понад 900 заявок не розглянуті.

Все це спеціаліст із найму робить вручну і, враховуючи кількість заявок, необхідно витратити декілька робочих днів лише на обробку відгуків.

З цього впливає третій зазначений пункт, а саме - зменшення вартості найму.

Згідно з даними Workable [39], нещодавні опитування показали, що середня ціна кожного найму для компанії становить трохи більше \$4,000. Тобто, якщо компанія планує найняти 100 співробітників на рік, то витрати на найм можуть становити понад \$400,000.

Звичайно, варто зважати, що Workable пропонує статистику Західного ринку і в Україні ці кошти будуть меншими, проте не настільки, щоб на них не зважати, адже крім стандартних публікацій вакансій та реклами, компанія має заплатити за кожну годину рекрутера, а також наймаючих менеджерів в компанії.

Враховуючи вищезазначене, в рамках дослідження буде розглядатися впровадження використання методів машинного навчання для такого етапу відбору як аналіз резюме кандидатів, що відправили відгук на позицію (рис. 1.23). Це надасть змогу зменшити навантаження, скоротить час, а також знизить необхідність залучати додаткових працівників, щоб обробляти всі можливі заявки. Та, як наслідок впровадження елементів ШІ, зменшить опосередковані витрати на оплату праці.

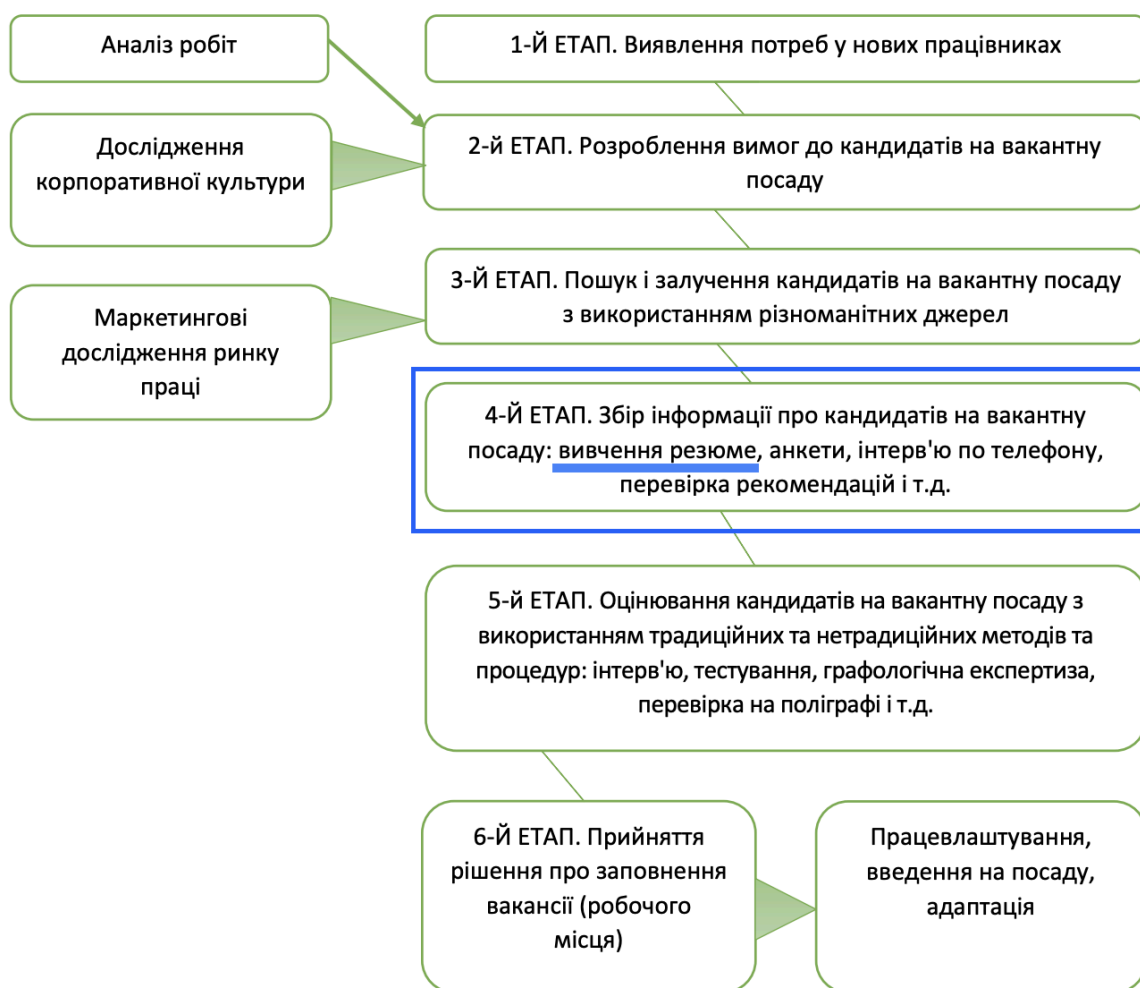


Рисунок 1.23 - Вибраний етап для автоматизації в класичному процесі відбору працівників [10]

## Висновки до Розділу 1

В розділі 1 було описано результати аналітичної роботи з літературними джерелами, що присвячені проблемі використання методів штучного інтелекту в сучасному рекрутингу. Кількість та різноманітність статей на цю тему свідчить про те, що ця тема активно набирає популярності протягом кількох останніх років і має широке поле для подальших досліджень та дискусій. Тема автоматизації відбору кандидатів, має як своїх прихильників, так і тих, хто з обережністю ставиться до впровадження використання штучного інтелекту в процесі найму. В тому числі, дискусійною є тема упередженості в прийнятті рішень, адже існує думка, що використання ШІ може приводити до неявної дискримінації кандидатів.

Також в цьому розділі було розглянуто деякі популярні ATS, що використовують в своїй роботі штучний інтелект. Окремо було проведене порівняння систем, популярних на українському ринку.

Окремим пунктом була розглянута цінність та доцільність використання методів автоматизованого аналізу резюме, де наведено статистичні дані про стан сучасного українського ІТ ринку, які свідчать про стрімкий ріст кількості кандидатів на позиції після початку повномасштабного вторгнення.

Підсумовуючи вищесказане, можна зробити висновок, що тема автоматизації обробки первинних відгуків покриває сучасні болі, з якими стикаються ІТ компанії при відборі кандидатів та дозволяє оптимізувати наявні ресурси.

Враховуючи наявну різноманітність методів, що можуть бути використані для автоматизації обробки резюме, в наступному розділі розглянемо та порівняємо деякі з них.

## 2 МЕТОДИ ВИРІШЕННЯ ЗАДАЧІ

### 2.1 Вибір методу обробки резюме

В останні роки відбувається стрімкий розвиток штучного інтелекту. Зараз існує відносно великий вибір методів, за допомогою яких можна працювати з текстом. Ці методи можуть визначати емоційну забарвленість тексту [40], шукати відповідності, розпізнавати спам, витягувати з тексту анотації, ранжувати дані (наприклад, резюме) в залежності від їхньої відповідності вказаним вимогам, виявляти закономірності та робити прогнози, аналізувати фотографії [41] тощо. Серед усіх цих варіантів необхідно обрати той, який буде найкраще справлятися з поставленою задачею та дозволить максимально реалізувати поставлену мету.

Під час аналізу наукових джерел, також були проаналізовані різні методи обробки резюме. Серед існуючої кількості різних методів, за допомогою яких можна працювати з текстом, було обрано, ті які є найбільш відповідними та оптимальними для використання в процесі аналізу резюме на етапі відбору резюме. Під час дослідження було обрано 5 методів, які спрямовані вирішити поставлені задачі дослідження. А саме:

- наївний байєсів класифікатор (Naive Bayes) [42];
- дерево рішень (Decision Tree) [42];
- метод опорних векторів (Support Vector Machines - SVM) [42];
- випадковий ліс (Random Forest) [43];
- KNN (K-Nearest Neighbors) [42].

Далі буде розглянуто та вивчено, обрані методи більш детально. Будуть розглянуті їхні особливості і чому вони можуть бути корисними у досягненні мети.

Наївний байєсів класифікатор (Naive Bayes) – це ймовірнісний метод класифікації, який застосовує теорему Байеса для обчислення ймовірності належності спостереження до певного класу, базуючись на наївному припущенні про незалежність змінних [44].

Математичне представлення Теорема Байеса [44]:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}, \quad (2.1)$$

де  $A$  і  $B$  є подіями,

$P(A)$  є апріорною ймовірністю гіпотези  $A$ ,

$P(B)$  є повною вірогідністю настання події  $B$ ,

$P(A | B)$  є ймовірністю гіпотези  $A$  за умови настання події  $B$ ,

$P(B | A)$  є ймовірністю настання події  $B$  за умови істинності гіпотези  $A$ .

Теорема Байеса дозволяє передбачити клас, використовуючи набір параметрів і ймовірність [44].

Спрощене рівняння класифікації даних визначає ймовірність належності до Class  $A$  на основі параметрів Feature 1 та Feature 2 за допомогою дроби та має вигляд [44]:

$$P(\text{Class } A | \text{Feature 1}, \text{Feature 2}) = \frac{P(\text{Feature 1} | \text{Class } A) * P(\text{Feature 2} | \text{Class } A) * P(\text{Class } A)}{P(\text{Feature 1}) * P(\text{Feature 2})} \quad (2.2)$$

де чисельник складається з добутку ймовірності того, що Feature 1 належить до Class  $A$ , ймовірності того, що Feature 2 належить до Class  $A$ , та ймовірності Class  $A$ ,

знаменник складається з добутку ймовірностей Ознаки 1 та Ознаки 2.

Naive Bayes може допомогти визначити ймовірність того, що резюме належить до певної категорії або має певні характеристики. Цей метод може бути ефективним для аналізу резюме, якщо важлива швидкість та легкість навчання моделі, а також коли є обмежена кількість даних. Він часто застосовується в ситуаціях, де текстові дані та категоризація є ключовими факторами.

Метод дерев рішень (decision trees) є одним з найпопулярніших підходів для вирішення задач класифікації та прогнозування. Цей метод іноді також називають деревами вирішальних правил, деревами класифікації або деревами регресії [42].

Алгоритми побудови дерев рішень включають два ключові етапи: "побудову" (tree building) і "скорочення" дерева (tree pruning). Під час побудови дерева приймаються рішення щодо вибору критерію для розщеплення вузлів і зупинки процесу навчання (якщо це передбачено алгоритмом). На етапі скорочення визначаються гілки, які потрібно відсікати для оптимізації моделі [42].

Процес побудови дерева здійснюється зверху вниз, тобто спадним чином. На цьому етапі важливо знайти критерій розщеплення, також відомий як критерій розбивки, який дозволяє розділити дані на підмножини, що асоціюються з певним вузлом перевірки. Кожен вузол перевірки позначається відповідним атрибутом [42].

Існує кілька критеріїв розщеплення, серед яких найпоширенішими є ентропія і індекс Джині (Gini). Ці критерії допомагають обрати атрибут на основі відстаней між розподілами класів [42].

Якщо дана множина  $T$ , що включає приклади з  $n$  класів, індекс Gini, тобто  $gini(T)$ , визначається за формулою [42]:

$$gini(T) = 1 - \sum_{j=1}^n p_j^2, \quad (2.3)$$

де  $T$  – поточний вузол,

$p_j$  – імовірність класу  $j$  у вузлі  $T$ ,

$n$  – кількість класів.

Тобто дерево рішень використовується для створення моделі, що дозволяє класифікувати або робити передбачення щодо конкретного об'єкту на основі декількох ознак. Він базується на створенні дерева, що допомагає зробити прогноз на основі порівняння вхідних даних з навчальною вибіркою.

Дерева прийняття рішень можуть бути корисним інструментом в контексті аналізу резюме, особливо коли потрібно враховувати багато критеріїв та приймати рішення на основі складних правил.

Метод опорних векторів (Support Vector Machine, SVM) належить до граничних методів, які визначають класи за допомогою кордонів областей. Цей метод використовується для розв'язання задач бінарної класифікації [42].

У SVM дані представляються в  $n$ -вимірному просторі, де  $n$  — кількість ознак. Класифікація виконується шляхом вибору відповідної гіперплощини, яка розділяє два класи. SVM знаходить таку гіперплощину, яка максимально віддаляє два класи один від одного, забезпечуючи правильне розділення навчальних даних. Для розділення лінійно нероздільних наборів даних SVM використовує функції ядра, які перетворюють початкові вектори даних у простір вищого виміру, також відомий як простір ознак. Опорні вектори — це вектори даних, які знаходяться на межах гіперплощини і визначають її положення та орієнтацію [45].

У методі опорних векторів класи позначаються як  $+1$  і  $-1$  (а не  $1$  і  $0$ ), а параметр зсуву явно позначається буквою  $b$  (а не включається у вектор  $w$  як константний доданок), що робить математичні викладки більш зрозумілими. Лінійний класифікатор описується наступною формулою [46]:

$$f(\bar{x}) = \text{sign}(\bar{w}^T \bar{x} + b), \quad (2.4)$$

де  $f(\bar{x})$  визначає, до якого класу належить вектор ознак,

$\bar{w}$  - ваги, які модель вивчає під час тренування,

$\bar{x}$  - вектор ознак вхідних даних,

$b$  - зсув (bias).

Значення  $-1$  позначає один клас, а  $+1$  — іншої.

Метод випадкового лісу є популярним розширенням методів дерев рішень для вирішення задач класифікації. Випадковий ліс - це ансамблевий метод машинного навчання, який використовується для класифікації, регресії та інших завдань. Він працює шляхом побудови багатьох дерев прийняття рішень під час тренування моделі та видає результат у вигляді моди класів (для класифікацій) або середнього прогнозу (для регресії) побудованих дерев [43].

Метод бегінга (bagging, або bootstrap aggregating) [47] був запропонований Лео Брейманом у 1994 році і є одним з перших ансамблевих методів у машинному навчанні. Основна ідея цього методу полягає в тому, що з навчальної вибірки  $X$  методом бутстрепа генеруються  $M$  вибірок  $X_1, \dots, X_M$ . На кожній вибірці навчається власний класифікатор  $a_i(X)$ . Підсумковий класифікатор формується як "середня" відповідь усіх класифікаторів (наприклад, шляхом голосування) [47]:

$$a(X) = \frac{1}{M} \sum_{i=1}^M a_i(X), \quad (2.5)$$

Метод Random Forest [47] є прикладом бегінга над деревами рішень і також використовує випадкові підпростори ознак. Основна ідея методу полягає в наступному.

Маємо навчальну вибірку  $X$  розміром  $l \times D$  (де  $l$  - кількість об'єктів, а  $D$  - кількість ознак) [47]:

- вибираємо кількість дерев у лісі:  $N$ ;
- кожне з  $N$  дерев будується на вибірці  $X_n$  ( $n = \overline{1, N}$ ), отриманій методом бутстрепа. При цьому дерево будується не на всьому ознаковому просторі, а лише з використанням  $d < D$  ознак, вибраних випадковим чином. Експериментальним шляхом отримані наступні рекомендації: у задачах класифікації вибирати  $d = \sqrt{D}$ , у задачах регресії  $d = D/3$ ;

- підсумковий класифікатор формується шляхом агрегування відповідей  $N$  дерев (для класифікації - вибір мажоритарного класу, для регресії - медіана або середнє значення).

KNN (K-Nearest Neighbors) [48] - метод «найближчого сусіда» або системи міркувань на основі аналогічних випадків. Відноситься до класу методів, робота яких ґрунтується на зберіганні даних у пам'яті для порівняння з новими елементами [48].

Метод найближчого сусіда (Nearest Neighbor – NN) [48, с. 7] є одним із найпростіших методів класифікації. Навчання в цьому методі полягає у

збереженні всіх об'єктів навчальної вибірки. Класифікація нового об'єкта  $X_0$  відбувається за наступним правилом (рис. 2.1): об'єкт відносять до того класу, представник якого знаходиться найближче до  $X_0$ . Формально це можна записати так:

так:

$$X_0 \in S_l : X_q \in S_l, \quad d(X_0, X_q) = \min d(X_0, X_i), \quad (2.6)$$

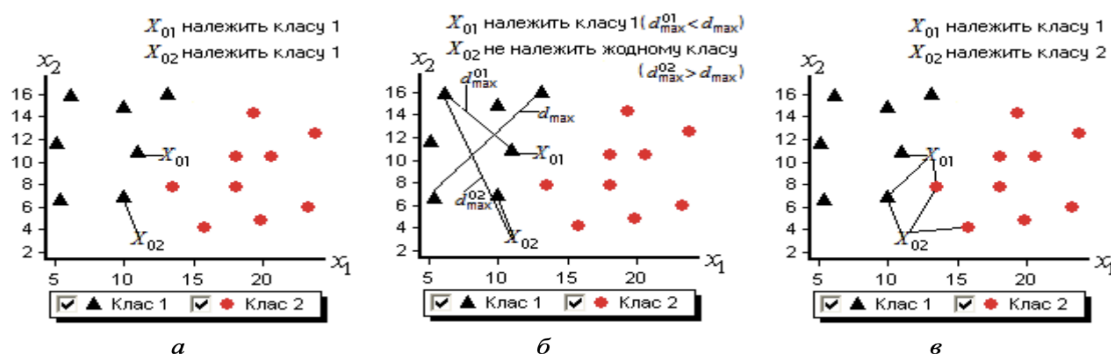


Рисунок 2.1 - Робота методів: а - ближнього сусіда; б - модифікації ближнього сусіда; в - к ближніх сусідів (k=3) [49, с.8].

Під час дослідження було обрано такі критерії, за якими було проведено порівняння:

- точність (Ассигасу): є фундаментальним показником ефективності класифікаційних моделей. Вона вказує на те, як вірно модель класифікує дані;
- швидкість: важлива для того, щоб прискорити процес розробки та експериментування з різними алгоритмами та параметрами;
- стійкість до перенавчання: важлива, щоб уникнути ситуацій, коли модель "запам'ятовує" тренувальні дані та не може адекватно узагальнити до нових;
- гнучкість та універсальність: дозволяють методу успішно застосовуватися до різноманітних видів завдань та даних, забезпечуючи більший спектр застосувань;

- стійкість до шуму: здатність працювати ефективно в умовах наявності шуму в даних. Наприклад, помилки введення даних або неправильне форматування, яке може виникнути під час збору даних.

## 2.2 Вирішення задачі багатокритеріального вибору

Результати оцінки за обраними критеріями, відображені в таблиці 2.1. Слід відмітити, що існує лише 2 показники, які можуть мати кількісну оцінку, це Точність та Швидкість навчання. Але ці критерії не мають сталих величин та будуть залежати від проведених експериментів та даних, на яких будуть апробуватися методи. Тому в ході дослідження було прийнято рішення оцінювати всі 5 критеріїв за якісною шкалою за допомогою аналізу інформаційних джерел щодо принципів роботи кожного з методів.

Таблиця 2.1 - Оцінка методів за обраними критеріями (якісна шкала)  
(Таблиця виконана самостійно)

	Точність (max)	Швидкість (max)	Стійкість до перенавчання (max)	Гнучкість та універсальність (max)	Стійкість до шуму (max)
Naive Bayes	Помірно висока	Висока	Висока	Залежить від умов	Висока
Дерево рішень	Залежить від параметрів та даних	Середня	Середня	Помірно висока	Середня
SVM	Висока	Залежить від параметрів та даних	Висока	Висока	Висока
Випадковий ліс	Висока	Висока	Залежить від параметрів та даних	Висока	Висока
KNN	Помірно висока	Залежить від параметрів та даних	Залежить від параметрів та даних	Залежить від параметрів та даних	Нижче середнього

Для здійснення розрахунків, була проведена класифікація та переведення значень у порядкову шкалу. Результати відображені у Таблиці 2.2:

- висока - 5 балів;
- помірно висока - 4 бали;
- середня/залежить від параметрів та даних - 3 бали;
- нижче середнього - 2 бали;
- низька - 1 бал.

Таблиця 2.2 - Оцінка методів за обраними критеріями (порядкова шкала)  
(Таблиця виконана самостійно)

	Точність	Швидкість навчання	Швидкість прогнозування	Стійкість до перенавчання	Гнучкість та універсальність	Стійкість до шуму
Naive Bayes	4	5	5	5	3	5
Дерево рішень	3	5	5	4	4	4
SVM	4	3	3	5	5	5
Випадковий ліс	5	4	4	4	5	5
KNN	4	3	3	3	3	2
Вагові коеф	(5)	(2)	(2)	(5)	(4)	(4)

В таблицях всі критерії відповідають твердження “чим вища оцінка, тим кращий показник”, тому приводити всі шкали до принципу оптимальності «за максимумом» немає необхідності.

Згідно з множиною Парето у критеріальній таблиці можуть опинитися альтернативи, які мають оцінки по всім критеріям гірші, ніж інші альтернативи.

Одразу зрозуміло, що такі альтернативи неконкурентоспроможні. Згідно з результатами оцінки - це Дерево рішень та KNN. Їх можна сміливо викреслювати з таблиці.

Нормалізуємо критерії та приведемо їх у шкалу від 0 до 1. Для цього скористаємося формулою [50, с. 14]:

$$f = \frac{f_{\text{вим}}}{f_{\text{ет}}}, \quad (2.7)$$

де  $f_{\text{вим}}$  - показник критерію,

$f_{\text{ет}}$  - еталонне значення.

За еталонне значення була обрана оцінка 5, як найбільший можливий показник.

Для дослідження було обрано лінійну згортку з ваговими коефіцієнтами, оскільки, по-перше, вона найбільш поширена, а отже зарекомендувала себе як якісний спосіб вимірювання, по-друге, обрані критерії мають різну вагу і значення при виборі методу автоматизованого аналізу резюме. Так, наприклад, вважаємо, що найвагомим критерієм є точність, оскільки маємо покладатися на роботу системи і не витратити час на постійні перевірки результатів. Безпосередній вплив на цей показник має стійкість до перенавчання, тому цьому значенню також присвоюється найвища вага - 5.

Трохи менше значення мають гнучкість та універсальність та стійкість до шуму, адже вони впливають на зручність та багатоваріантність використання методу, та здатність методу працювати навіть з “зашумленими” даними відповідно. Тому присвоюємо цим показникам ваговий коефіцієнт 4.

І найменш значущим критерієм в рамках дослідження є швидкість, оскільки планується відносно невелика та обмежена кількість даних, а також є допустимим певний час на обробку заради точності результату. Тому цьому критерію призначено значення 2.

Далі було розраховано вагові коефіцієнти методом простого ранжування. Результати наведено у таблиці 2.3.

Таблиця 2.3 - Оцінка методів за обраними критеріями (нормалізовані критерії) (Таблиця виконана самостійно)

	Точність	Швидкість навчання	Швидкість прогнозування	Стійкість до перенавчання	Гнучкість та універсальність	Стійкість до шуму
Naive Bayes	0.8	1	1	1	0.6	1
SVM	0.8	0.6	0.6	1	1	1
Випадковий ліс	1	0.8	0.8	0.8	1	1
Вагові коеф	0.23	0.09	0.09	0.23	0.18	0.18

Формула лінійної адитивної згортки з ваговими коефіцієнтами [51, с. 44]:

$$Z = \max \sum_{j=1}^n \alpha_j \beta_j, \quad (2.8)$$

де  $Z$  - узагальнений показник (результат згортки),

$\alpha_j$  - значення  $i$ -ої складової,

$\beta_j$  - ваговий коефіцієнт  $i$ -ої складової,

$n$  - кількість складових.

Проведемо розрахунки за формулою.

Для Naive Bayes:  $0.8*0.23+1*0.09+1*0.09+1*0.23+0.6*0.18+1*0.18 =$   
 $= 0.184 + 0.09 + 0.09 + 0.23 + 0.108 + 0.18 = 0.882$

SVM:  $0.184 + 0.054 + 0.054 + 0.23 + 0.18 + 0.18 = 0.882$

Випадковий ліс:  $0.23 + 0.072 + 0.072 + 0.184 + 0.18 + 0.18 = 0.918$

Згідно з розрахунками найбільш підходящим методом для використання в

рамках автоматизованого відбору резюме є Випадковий ліс.

### Висновки до Розділу 2

У Розділі 2 було описано та математично представлено п'ять методів, що можуть використовуватися для автоматизованого аналізу резюме.

Також, в ході вирішення задачі багатокритеріального вибору, було обрано альтернативи та критерії їхньої оцінки. Всі критерії були нормалізовані та приведені у шкалу від 0 до 1. Ці дані використовувалися для розрахунку корисності альтернатив за лінійною адитивною згорткою з ваговими коефіцієнтами.

За результатами проведених розрахунків, можна зробити висновок, що найкращою альтернативою є використання Випадкового лісу, тоді як Дерево рішень та KNN, відповідно до множини Парето, можна було не розглядати, адже вони по всім критеріям гірші, ніж інші альтернативи.

### 3 ПРОВЕДЕННЯ ЕКСПЕРИМЕНТУ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ

#### 3.1 Визначення вхідних даних та опис алгоритму

Для того, щоб належним чином дослідити можливості автоматизованого відбору резюме за допомогою методів машинного навчання, необхідно відібрати відповідні дані для тренування та тестування.

Оскільки існує проблема з обробкою української мови через непопулярність її використання в ІТ сфері [52], з метою проведення експерименту було обрано англomовний набір даних [53], що містить дані про 70+ тисяч кандидатів.

Датасет містить наступні стовпці: №, Age, Accessibility, EdLevel, Employment, Gender, MentalHealth, MainBranch, YearsCode, YearsCodePro, Country, PreviousSalary, HaveWorkedWith, ComputerSkills, Employed.

Під час проведення експерименту використовувалася мова програмування Python та її бібліотеки, в тому числі Pandas, Jupiter Notebook.

Для проведення експерименту немає необхідності використовувати всі наявні дані з оригінального датасету, тому вони були відфільтровані. Під час дослідження було використано наступні дані: Employment, MainBranch, YearsCode, PreviousSalary, HaveWorkedWith та Employed. З них Employment, YearsCode, PreviousSalary та Employed мають чисельні значення, а MainBranch та ComputerSkills є рядками типу String. Оскільки більшість алгоритмів машинного навчання працюють з числовими даними, для того, щоб можна було використовувати ці ознаки у моделі навчання, було використано кодування категоріальних ознак 'LabelEncoder' (для перетворення категоріальних значень у числові, які можна подавати на вхід у модель машинного навчання). LabelEncoder присвоює унікальний цілочисельний ідентифікатор кожному унікальному значенню в категоріальному стовпці (рис. 3.1).

```
# Застосування Label Encoding до категоріальних ознак
label_encoder = LabelEncoder()
X['HaveWorkedWith'] = label_encoder.fit_transform(X['HaveWorkedWith'])
X['MainBranch'] = label_encoder.fit_transform(X['MainBranch'])
```

Рисунок 3.1 - Перетворення категоріальних значень у числові (Рисунок виконаний самостійно)

Далі, для зручності та читабельності коду була створена функція, для навчання та оцінки моделі (рис. 3.2).

```
# Функція для навчання і оцінки моделі
def evaluate_model(X_train, y_train, X_test, y_test, add_noise=False):
    if add_noise:
        noise = np.random.normal(loc=0, scale=1, size=len(X_train))
        X_train = X_train.copy()
        X_train['HaveWorkedWith'] = X_train['HaveWorkedWith'] + noise

    # Навчання моделі
    start_time = time.time()
    model = SVC(random_state=0)
    model.fit(X_train, y_train)
    training_time = time.time() - start_time

    # Крос-валідація
    cv_scores = cross_val_score(model, X_train, y_train, cv=5)
    cv_mean = cv_scores.mean()

    # Прогнозування
    start_time = time.time()
    y_pred = model.predict(X_test)
    prediction_time = time.time() - start_time

    # Оцінка точності та F1-оцінки
    accuracy = accuracy_score(y_test, y_pred)
    f1 = f1_score(y_test, y_pred)

    # Матриця плутанини
    conf_matrix = confusion_matrix(y_test, y_pred)

    return training_time, prediction_time, accuracy, f1, cv_mean, conf_matrix
```

Рисунок 3.2 - Функція для навчання і оцінки моделі (на прикладі SVM) (Рисунок виконаний самостійно)

Щоб оцінити точність побудованих моделей на основі обраних методів, була використана метрика 'accuracy\_score' (рис 3.3). 'accuracy\_score' - це метрика, яка використовується для оцінки точності моделі класифікації. Вона вимірює частку правильно класифікованих екземплярів серед усіх екземплярів у тестовому наборі.

З її допомогою можна оцінити, наскільки добре працює модель в цілому, а також використовувати її для порівняння точності різних методів на тестовому наборі даних.

```
# Оцінка точності моделі
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print("Точність моделі на тестових даних:", accuracy)
```

Рисунок 3.3 - Оцінка точності моделі навчання (Рисунок виконаний самостійно)

Оцінка швидкості моделей навчання була проведена за допомогою функції ‘time.time()’ (рис. 3.4, 3.5, 3.6 для NaiveBayes, SVM та random Forest відповідно).

```
# Навчання моделі
start_time = time.time()
model = GaussianNB()
model.fit(X_train, y_train)
training_time = time.time() - start_time
```

Рисунок 3.4 - Оцінка швидкості моделі навчання Naive Bayes (Рисунок виконаний самостійно)

```
# Навчання моделі
start_time = time.time()
model = SVC(random_state=0)
model.fit(X_train, y_train)
training_time = time.time() - start_time
```

Рисунок 3.5 - Оцінка швидкості моделі навчання SVM (Рисунок виконаний самостійно)

```
# Навчання моделі
start_time = time.time()
model = RandomForestClassifier(random_state=0)
model.fit(X_train, y_train)
training_time = time.time() - start_time
```

Рисунок 3.6 - Оцінка швидкості моделі навчання Random Forest (Рисунок виконаний самостійно)

Крім швидкості навчання також було проведено оцінку швидкості прогнозування (рис. 3.7).

```
# Прогнозування
start_time = time.time()
y_pred = model.predict(X_test)
prediction_time = time.time() - start_time
```

Рисунок 3.7 - Оцінка швидкості прогнозування (Рисунок виконаний самостійно)

Оцінка стійкості до перенавчання моделей була проведена за допомогою крос-валідації (рис. 3.8). Цей спосіб допомагає виявити, чи має модель навчання тенденцію до перенавчання (тобто, чи модель надто адаптована до тренувальних даних та чи може ефективно узагальнювати на нові дані). Це важливо для того, щоб мати впевненість у тому, що модель працюватиме на реальних даних так само добре, як і на тренувальних.

```
# Оцінка стійкості моделі до перенавчання за допомогою крос-валідації
cv_scores = cross_val_score(model, X_train, y_train, cv=5)
print("Оцінка стійкості моделі до перенавчання (середнє по 5 фолдам):", cv_scores.mean())
```

Рисунок 3.8 - Оцінка стійкості моделі до перенавчання (Рисунок виконаний самостійно)

Для оцінки стійкості до шуму, був згенерований випадковий шум за допомогою `np.random.normal`, з параметрами `loc=0`, `scale=1`, та розміром, що відповідає кількості рядків у навчальному наборі (рис 3.9). Шум було додано до

стовпчика 'HaveWorkedWith' у навчальному наборі. Після чого було проведено вимірювання точності та швидкості моделі навчання, але вже на зашумлених даних. Порівняння результатів з початковою моделлю показує, чи змінилися результати та наскільки і дозволяє, таким чином, оцінити, чи є модель стійкою до шуму.

```
if add_noise:
    noise = np.random.normal(loc=0, scale=1, size=len(X_train))
    X_train = X_train.copy()
    X_train['HaveWorkedWith'] = X_train['HaveWorkedWith'] + noise
```

Рисунок 3.9 - Додавання шуму до даних (Рисунок виконаний самостійно)

Застосування F1-оцінки та аналізу матриці помилок допомагає в оцінці гнучкості та універсальності моделі, оскільки ці метрики дають змогу зрозуміти, як модель веде себе в різних умовах і з різними типами даних, а також допомагають ідентифікувати можливі напрямки для поліпшення моделі (рис 3.10).

```
# F1-оцінка моделі
f1 = f1_score(y_test, y_pred)
print("F1-оцінка моделі на тестових даних:", f1)

F1-оцінка моделі на тестових даних: 0.7003484320557491

# Матриця плутанини
conf_matrix = confusion_matrix(y_test, y_pred)
print("Матриця плутанини:")
print(conf_matrix)

Матриця плутанини:
[[4938 3529]
 [2663 7236]]
```

Рисунок 3.10 - Застосування F1 та матриці помилок (Рисунок виконаний самостійно)

В даному випадку 'f1\_score()' розраховує F1-оцінку, яка бере до уваги і точність, і повноту, забезпечуючи баланс між ними. 'confusion\_matrix()' подає матрицю помилок, яка відображає детальніше, де і як модель робить помилки.

Якщо модель навчання показує однакову продуктивність на звичайних та зашумлених даних, свідчить про високу гнучкість та універсальність. Це означає,

що модель здатна ефективно обробляти вхідні дані, навіть коли вони включають статистичний шум, що може імітувати реальні умови, де дані часто не ідеальні. Втім, для повного висновку про універсальність бажано провести тестування моделі на різних типах задач та даних, включаючи різні домени або різні розподіли даних.

Загальна блок-схема алгоритму виглядає наступним чином (рис.3.7) :

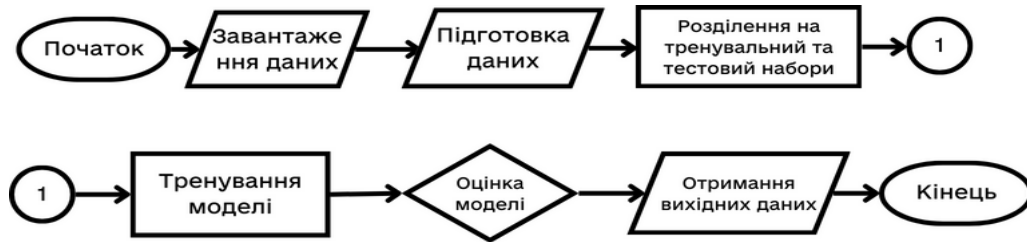


Рисунок 3.7 - Блок-схема алгоритму оцінки моделей (Рисунок виконаний самостійно)

### 3.2 Аналіз отриманих результатів

Під час проведення експерименту, кожен з методів був виконаний по 20 разів, з метою простеження закономірностей та отримання більш точних результатів. У таблицях 3.1 - 3.3 наведено результати експериментів з дослідження точності, швидкості стійкості до перенавчання та стійкості до шуму.

Таблиця 3.1 - Показники моделі Naive Bayes (Таблиця виконана самостійно)

Показник	Мін значення (сек)	Середнє значення (сек)	Макс значення (сек)	Мін значення (сек) + шум	Середнє значення (сек) + шум	Макс значення (сек) + шум
Час навчання моделі	0.0075	0.0081	0.0140	0.0075	0.0084	0.0131
Час прогнозування на тестових даних	0.0013	0.0019	0.0051	0.0012	0.0014	0.0026
Оцінка стійкості моделі до перенавчання	0.6621	0.6633	0.6681	0.6621	0.6648	0.6682
Точність моделі	0.6572	0.6598	0.6714	0.6572	0.6629	0.6714



Згідно з даними з Таблиці 3.1. можна бачити наступне: показники швидкості моделі Naive Bayes доволі високі і не мають суттєвих відмінностей між оригінальними та зашумленими даними.

Якщо говорити про такі показники як стійкість до перенавчання та точність моделі, то ці показники тим кращі, чим їхні значення ближче до 1. Враховуючи, що середня точність моделі дорівнює 0.66 та важливість цього критерію в ситуації обробки резюме, то можна говорити скоріше про середній результат. Те саме можна сказати і про оцінку стійкості до перенавчання.

Трохи кращі показники має модель навчання на основі методу Випадкового лісу. Хоча і має нижчу швидкість, проте показники стійкості моделі до перенавчання та точності дорівнюють 0.7384 та 0.7412 відповідно. Цей результат вже можна назвати вище середнього, особливо у порівнянні з моделлю Naive Bayes. Вагомої різниці між оригінальними та зашумленими даними також не спостерігається

Аналізуючи результати роботи SVM моделі, можна прийти до висновку, що з наявним датасетом цей алгоритм впорався найгірше. Модель навчання на основі методу SVM, як і попередні, показала гарну стійкість до шуму, але виявилася надзвичайно повільною (середній показник для навчання моделі становить 48.7160 секунд та 20.5404 секунд для прогнозування на тестових даних). Стійкість до перенавчання та точність моделі виявилися середніми та схожі з показниками Naive Bayes.

Далі було проведено порівняння та вирішення задачі багатокритеріального вибору вже на основі отриманих експериментальних даних. Розраховуємо результати за формулою (2.8), що використовувалася у попередніх розділах.

$$\text{Naive Bayes: } 0.6598*0.23 + 3*0.09 + 3*0.09 + 0.6633*0.23 + 2*0.18 + 3*0.18 = 1.744$$

$$\text{SVM} = 0.6548*0.23 + 1*0.09 + 1*0.09 + 0.6613*0.23 + 2*0.18 + 3*0.18 = 1.383$$

$$\text{Випадковий ліс} = 0.7412*0.23 + 2*0.09 + 2*0.09 + 0.7384*0.23 + 3*0.18 + 3*0.18 = 1.78$$

Таблиця 3.4 - Оцінка методів за обраними критеріями (Таблиця виконана самостійно)

	Точність	Швидкість навчання	Швидкість прогнозування	Стійкість до перенавчання	Гнучкість та універсальність	Стійкість до шуму
Naive Bayes	0.6598	3	3	0.6633	2	3
SVM	0.6548	1	1	0.6613	2	3
Випадковий ліс	0.7412	2	2	0.7384	3	3
Вагові коеф	0.23	0.09	0.09	0.23	0.18	0.18

Таким чином, можна підтвердити, що найкращим варіантом для роботи з автоматизацією відбору резюме є метод Випадкового лісу.

### Висновок до Розділу 3

В розділі 3 було описано датасет, який використовувався для моделей навчання на основі методів Naive Bayes, SVM та Random Forest, а також описана блок-схема та сама реалізація цих методів машинного навчання.

Дані, що були отримані в ході експериментів були використані для порівняння методів та вибору найефективнішого з них, а саме Random Forest. В подальшому цей метод можна використовувати для вдосконалення ATS шляхом розробки бібліотеки, яка автоматизує процес оцінки та ранжування кандидатів на основі їхньої відповідності вакансіям. Використання результатів моделювання для розробки такої бібліотеки може значно покращити ефективність та точність системи.

## ВИСНОВКИ

У результаті роботи над дослідженням було проаналізовано актуальність та доцільність використання методів штучного інтелекту для підвищення ефективності в процесі рекрутингу. Для цього було виконано наступні задачі:

- було розглянуто сучасний стан ІТ ринку в Україні, а також наукові публікації та системи відстеження кандидатів, популярні в країні та світі;
- за допомогою статистичних даних, було досліджено цінність використання машинного навчання у рекрутингових системах;
- було визначено популярні методи, які можуть використовуватися для автоматизованого аналізу резюме, а також проведено їхнє порівняння;
- визначено та використано метрики для оцінки ефективності застосування обраних методів;
- обрано оптимальний метод для автоматизованого аналізу резюме за допомогою вирішення задачі багатокритеріального вибору.

В ході дослідження було розглянуто різні методи автоматизованого аналізу резюме в рекрутингових системах. Для цього було наведено порівняння п'яти популярних методів, за допомогою яких можна порівнювати та класифікувати дані, а саме:

- наївний басієв класифікатор (Naive Bayes);
- дерево рішень (Decision Tree);
- опорних векторів (Support Vector Machines - SVM);
- випадковий ліс (Random Forest);
- KNN (K-Nearest Neighbors).

Для порівняння цих методів було обрано та використано 6 критеріїв: точність, швидкість навчання, швидкість прогнозування, стійкість до перенавчання, гнучкість та універсальність та стійкість до шуму.

Порівняння методів було проведено за допомогою вирішення задачі багатокритеріального вибору.

За результатами проведеного дослідження найкращим варіантом виявився метод Випадкового лісу. Проте, у цьому напрямку є велике поле для подальших досліджень, наприклад, проведення експериментів з різними наборами даних та вдосконалення самої моделі, адже, хоча були отримані результати вище середнього, при автоматизації відбору резюме точність роботи алгоритму дуже важлива. Можливі помилки при відборі навпаки можуть погіршити якість роботи рекрутерів, призвести до упередженості, втратити найкращих кандидатів та підвищити вартість найму.

В подальшому цей метод можна використовувати для вдосконалення ATS шляхом розробки бібліотеки, яка автоматизує процес оцінки та ранжування кандидатів на основі їхньої відповідності вакансіям.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Таран А. О. ЗАСТОСУВАННЯ АІ В РЕКРУТИНГОВИХ СИСТЕМАХ 28-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті». Зб. матеріалів форуму. Т. 6., – Харків: ХНУРЕ. 2024. – 846-847.
2. Oswal N., Ateeq K., Mathew S. Trends in Recruitment Information and Communication System using Artificial Intelligence in Industry 4.0. In: FEMIB. 2021. p. 111-118.
3. Mariani K., Vega-Lozada, F. The Use of AI and Algorithms for Decision-making in Workplace Recruitment Practices. *Journal of Student Research*, 2023, 12.1.
4. El Mohadab M., Bouikhalene B., Safi, S. Automatic CV processing for scientific research using data mining algorithm. *Journal of King Saud University-Computer and Information Sciences*, 2020, 32(5), 561-567.
5. Гей А. В. Спосіб оцінки відповідності кандидата до вакансії на базі засобів машинного навчання : магістерська дис. : 123 Комп'ютерна інженерія / Гей Антон Віталійович. – Київ, 2024. – 81 с.
6. Apatean A., Szakacs E., Tilca M. Machine-learning based application for staff recruiting. *Acta Technica Napocensis*, 2017, 58.4: 16-21.
7. Alfawareh H., Jusoh, S. Intelligent decision support system for CV evaluation based on natural language processing. *International Journal of Advanced and Applied Sciences*, 2019, 6.4: 1-8.
8. Faliagka E., Ramantas K., Tsakalidis A., Tzimas G. Application of machine learning algorithms to an online recruitment system. In: *Proc. International Conference on Internet and Web Applications and Services*. 2012. p. 215-220.
9. Талько Ю. С. Інформаційна система пошуку та відбору персоналу з використанням нейронних мереж. MS thesis. Київ, 2018.
10. УПРАВЛІННЯ ПРОЦЕСОМ РЕКРУТИНГУ НА ПІДПРИЄМСТВАХ В СУЧАСНИХ УМОВАХ. URL: [http://www.economy.nayka.com.ua/pdf/1\\_2018/47.pdf](http://www.economy.nayka.com.ua/pdf/1_2018/47.pdf) (дата останнього звернення 12.05.2024).

11. Каушан С. О. Метод та засоби інформаційної системи електронного рекрутингу на основі машинного навчання : кваліфікаційна робота магістра : 126 Інформаційні системи та технології / С. О. Каушан ; Хмельниц. нац. ун-т. – Хмельницький, 2023. – 119 с.
12. Осадча, К. П., Осадчий В. В., Харитонов О. А. Підбір персоналу підприємств з використанням автоматизованих систем. *Ukrainian Journal of Educational Studies and Information Technology* 9.2 (2021): 1-19.
13. Вонберг Т. В., Головка А. А. Рекрутинг персоналу в епоху діджиталізації. *Бізнес Інформ.* 2020. №6. С. 313–318. <https://doi.org/10.32983/2222-4459-2020-6-313-318>
14. Кравчук О. І., Варіс І. О., Бідна Т. О.. Цифрові технології рекрутингу персоналу. *Маркетинг і цифрові технології* 6.1 (2022): 92-110.
15. Дьогтева, І. О., and Л. І. Рудь. Автоматизація HR-процесів в HRM-системах. Diss. ВНТУ, 2021.
16. Applicant Tracking System, ATS. URL: <https://www.it.ua/knowledge-base/technology-innovation/applicant-tracking-system-ats> (дата останнього звернення 12.05.2024).
17. Köchling A., Marius C. W. Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research* 13.3 (2020): 795-848. URL: <https://link.springer.com/article/10.1007/s40685-020-00134-w> (дата останнього звернення 02.06.2024)
18. Hunkenschroer A. L., Luetge C. Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, 2022, 178.4: 977-1007.
19. Sharonova, N., Kyrychenko, I., Gruzdo, I., Tereshchenko, G. Generalized Semantic Analysis Algorithm of Natural Language Texts for Various Functional Style Types, in *CEUR Workshop Proceedings*, 2022, 3171, pp. 16-26.

20. Grunenberg E, Peters H, Francis MJ, Back MD, Matz SC Machine learning in recruiting: predicting personality from CVs and short text responses. *Front. Soc. Psychol.* 1:1290295, 2024. doi: 10.3389/frsps.2023.1290295.
21. Tiwari V., Sapna J. C. Intelligent Hiring with Resume Parser and Ranking Using Machine Learning and Natural Language Processing. URL: <https://ijrsc.in/pdf/Versha011224.pdf> (дата останнього звернення 12.05.2024).
22. Strohmeier S., Piazza, F. Artificial intelligence techniques in human resource management - a conceptual exploration. *Intelligent Techniques in Engineering Management: Theory and Applications*, 2015, 149-172.
23. Smelyakov K., Hurova Y., Osiiievskiy S. Analysis of the Effectiveness of Using Machine Learning Algorithms to Make Hiring Decisions, 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2023), April 20–21, 2023. – CEUR-WS, 2023, ISSN 1613-0073. – Volume 3387, PP. 77-92.
24. 11 Best Applicant Tracking Systems Of 2023. URL: <https://www.forbes.com/advisor/business/best-applicant-tracking-systems/> (дата останнього звернення 20.04.2024).
25. The 16 Best Applicant Tracking Systems (ATS) in Q1 2024. URL: <https://www.selectsoftwarereviews.com/buyer-guide/applicant-tracking-systems#introduction-ver2> (дата останнього звернення 20.04.2024).
26. BreezyHR. URL: <https://breezy.hr/> (дата останнього звернення 20.04.2024).
27. Manatal. URL: <https://www.manatal.com/> (дата останнього звернення 20.04.2024).
28. GreenHouse. URL: <https://www.greenhouse.com/uk> (дата останнього звернення 20.04.2024).
29. Avature. URL: <https://www.avature.net/applicant-tracking-system/> (дата останнього звернення 20.04.2024).
30. SmartRecruiters. URL: <https://www.smartrecruiters.com/> (дата останнього звернення 20.04.2024).

31. iCIMS. URL: <https://www.icims.com/> (дата останнього звернення 20.04.2024).

32. Zoho Recruit. URL: <https://www.zoho.com/recruit/> (дата останнього звернення 20.04.2024).

33. 5 популярних систем управління кандидатами (ATS): огляд для рекрутерів. URL: <https://happymonday.ua/5-populyarnyh-ats> (дата останнього звернення 20.04.2024).

34. CleverStaff. URL: [https://cleverstaff.net/ua/?utm\\_source=google&utm\\_medium=cpc&utm\\_campaign=ua\\_srch\\_brand\\_main&gad\\_source=1&gclid=Cj0KCCQiA7aSsBhCiARIsALFvovwdFVF6dBfSI9UWN-2mq4dqH7wr6xiHb\\_2BE8IToYrMIT0uz8t0uGAaAt0zEALw\\_wcB](https://cleverstaff.net/ua/?utm_source=google&utm_medium=cpc&utm_campaign=ua_srch_brand_main&gad_source=1&gclid=Cj0KCCQiA7aSsBhCiARIsALFvovwdFVF6dBfSI9UWN-2mq4dqH7wr6xiHb_2BE8IToYrMIT0uz8t0uGAaAt0zEALw_wcB) (дата останнього звернення 21.04.2024).

35. Hurma. URL: <https://hurma.work/> (дата останнього звернення 21.04.2024).

36. Workable. URL: <https://www.workable.com/> (дата останнього звернення 20.04.2024).

37. Формуємо ціннісну пропозицію і бізнес-модель для культурної спадщини. URL: <https://reherit.org.ua/formuyemo-tsinnisnu-propozytsiyu-i-biznes-model-dlya-kulturnoy-i-spadshhyny/> (дата останнього звернення 27.04.2024).

38. Аналітика Djinni. URL: [https://blog.djinni.co/post/2023-report?utm\\_source=website&utm\\_medium=banner&utm\\_campaign=2023-report](https://blog.djinni.co/post/2023-report?utm_source=website&utm_medium=banner&utm_campaign=2023-report) (дата останнього звернення 20.05.2024).

39. Recruiting costs FAQ: Budget and cost per hire. URL: <https://resources.workable.com/tutorial/faq-recruitment-budget-metrics> (дата останнього звернення 20.04.2024).

40. Nazarenko D., Afanasieva I., Golian N., Golian V. Investigation of the deep learning approaches to classify emotions in texts CEUR Workshop Proceedings, 2021, 2870, стр. 206–224.

41. Smelyakov K., Chupryna A., Bohomolov O., Hunko N. The Neural Network Models Effectiveness for Face Detection and Face Recognition, 2021 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), 2021, pp. 1-7, doi: 10.1109/eStream53087.2021.9431476.

42. Тема 9. Методи дерев рішень, класифікації та прогнозування. URL: [https://moodle.znu.edu.ua/pluginfile.php?file=/486136/mod\\_resource/content/1/%D0%9B%D0%B5%D0%BA%D1%86%D1%96%D1%8F%209.pdf](https://moodle.znu.edu.ua/pluginfile.php?file=/486136/mod_resource/content/1/%D0%9B%D0%B5%D0%BA%D1%86%D1%96%D1%8F%209.pdf) (дата останнього звернення 25.05.2024).

43. Random forest URL: [https://uk.wikipedia.org/wiki/Random\\_forest](https://uk.wikipedia.org/wiki/Random_forest) (дата останнього звернення 25.05.2024).

44. Методичні вказівка до лабораторних робіт. Інтелектуальний аналіз даних. ХНУРЕ. URL: [https://drive.google.com/file/d/1dmNREY7ESGE0SXi0fdQ5r04c8AhI\\_App/view?usp=sharing](https://drive.google.com/file/d/1dmNREY7ESGE0SXi0fdQ5r04c8AhI_App/view?usp=sharing) (дата останнього звернення 25.05.2024).

45. Системи Data Science. Практикум з системної інженерії систем Data Science/КПІ ім. І. Сікорського. Київ, - 2022. - 67с. URL: [https://ela.kpi.ua/bitstream/123456789/50010/4/Data\\_Science\\_2022\\_23.pdf](https://ela.kpi.ua/bitstream/123456789/50010/4/Data_Science_2022_23.pdf) (дата останнього звернення 25.05.2024).

46. Метод опорних векторів. Лекція 7. URL: [http://om.univ.kiev.ua/users\\_upload/15/upload/file/pr\\_lecture\\_07.pdf](http://om.univ.kiev.ua/users_upload/15/upload/file/pr_lecture_07.pdf) (дата останнього звернення 25.05.2024).

47. Decision Trees and Random Forest (UA). URL: <https://www.kaggle.com/code/emstrakhov/decision-trees-and-random-forest-ua> (дата останнього звернення 25.05.2024).

48. Метод "найближчого сусіда" або системи міркувань на основі аналогічних випадків. URL: <https://studfile.net/preview/7818687/page:2/> (дата останнього звернення 25.05.2024).

49. Мацуга, О.М. Навчальний посібник до вивчення курсу «Інформаційні технології розпізнавання образів» [Текст] / О.М. Мацуга, Ю.М. Архангельська, Н.М. Єрещенко. – Д.: ПВВ ДНУ, 2016. – 60 с.

50. Теорія ігор та прийняття рішень. Лекція 6: Введення до теорії корисності / О. Мазурова. ХНУРЕ. URL:

[https://dl.nure.ua/pluginfile.php/793956/mod\\_resource/content/6/%D0%A2%D0%86%D1%82%D0%B0%D0%9F%D0%A0\\_%D0%BB%D0%BA\\_6.pdf](https://dl.nure.ua/pluginfile.php/793956/mod_resource/content/6/%D0%A2%D0%86%D1%82%D0%B0%D0%9F%D0%A0_%D0%BB%D0%BA_6.pdf) (дата останнього звернення 27.05.2024).

51. Теорія ігор та прийняття рішень. Лекція 5: Задачі багатокритеріального прийняття рішень / О. Мазурова. ХНУРЕ. URL:

[https://dl.nure.ua/pluginfile.php/793954/mod\\_resource/content/6/%D0%A2%D0%86%D1%82%D0%B0%D0%9F%D0%A0\\_%D0%BB%D0%BA\\_5.pdf](https://dl.nure.ua/pluginfile.php/793954/mod_resource/content/6/%D0%A2%D0%86%D1%82%D0%B0%D0%9F%D0%A0_%D0%BB%D0%BA_5.pdf) (дата останнього звернення 27.05.2024).

52. Panchenko D., Maksymenko D., Turuta O., Yerokhin A., Daniil Y., Turuta O. Evaluation and Analysis of the NLP Model Zoo for Ukrainian Text Classification // Communications in Computer and Information Science. 2022. P. 109–123.

53. 70k+ Job Applicants Data (Human Resource). URL: <https://www.kaggle.com/datasets/ayushtankha/70k-job-applicants-data-human-resource/data> (дата останнього звернення 30.05.2024)