

АНОТАЦІЯ

Міньківська М.В. Дослідження ефективності систем пошуку роботи методами **Data Science**. Спеціальність 122 «Комп'ютерні науки», Освітня програма «Комп'ютерні технології обробки даних (Data Science)». Донецький національний університет імені Василя Стуса, Вінниця, 2022.

Магістерська робота присвячена дослідження ефективності систем пошуку роботи методами **Data Science**. Об'єктом дослідження даної роботи є процес оцінювання ефективності систем пошуку роботи.

Предметом дослідження даної роботи є методи **Data Science** для аналізу статистичних даних, які можуть використовуватися для оцінювання ефективності систем пошуку роботи.

Встановлений перелік факторів що впливає на ефективність працевлаштування. Розроблено моделі і здійснена програмна реалізація прогнозу ефективності працевлаштування за можливими наборами факторів, що впливають на ефективність працевлаштування.

Ключові слова: система пошуку роботи, ефективність працевлаштування, методи **Data Science**.

68 с., 2 табл., 34 рис., 1 дод. 51 джерел.

ABSTRACT

Minkivska M.V. Efficiency research of job search systems by **Data Science methods**. Specialty 122 "Computer science", educational program "Computer technologies of data processing (Data Science)". **Vasyl' Stus Donetsk National University**, Vinnytsia, 2022.

The master's thesis is devoted to the study of the effectiveness of the job search system using **Data Science** methods. The object of research of this work is the process of evaluating the effectiveness of the job search system.

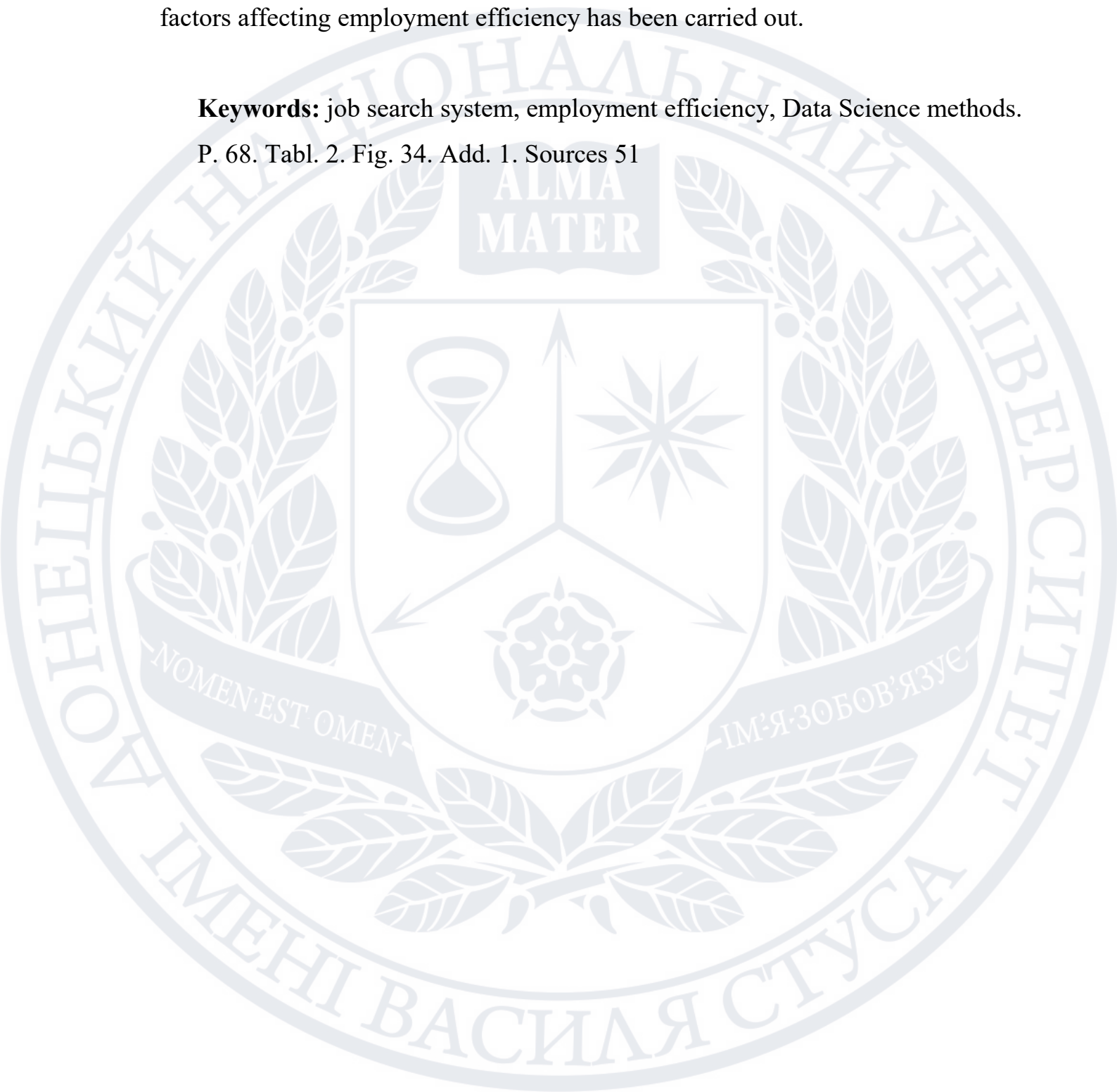
The subject of research in this work is **Data Science** methods for analyzing statistical data that can be used to evaluate the effectiveness of the job search system.

The indicators that characterize employment for a job candidate and an employer

in an automated job search system are defined. A list of factors affecting employment efficiency is established. Models have been developed and software implementation of the forecast of employment efficiency based on possible sets of factors affecting employment efficiency has been carried out.

Keywords: job search system, employment efficiency, Data Science methods.

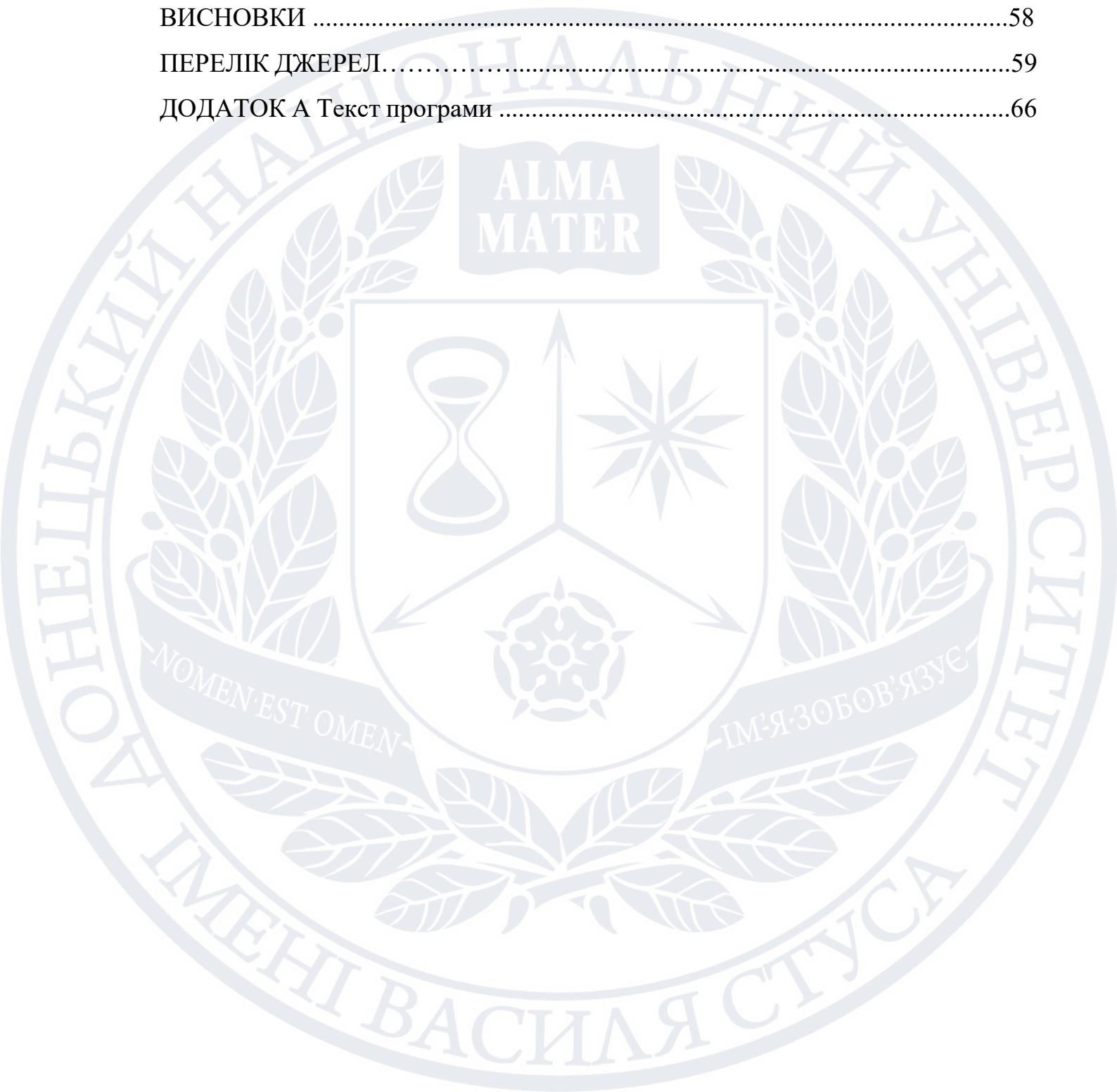
P. 68. Tabl. 2. Fig. 34. Add. 1. Sources 51



ЗМІСТ

ВСТУП	6
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ.....	8
1.1 Дослідження існуючих систем пошуку роботи за цільовою аудиторією, функціоналом, характеристиками наборів даних про вакансії і претендентів.....	8
1.2. Дослідження реалізації зворотного зв'язку щодо ефективності працевлаштування. Методи оцінки ефективності систем пошуку роботи.....	10
1.3. Дослідження методів обробки даних в системах пошуку роботи.....	14
1.4 Постановка задачі (розробки моделей для інтелектуальної функції систем пошуку роботи, що враховує ефективність працевлаштування).....	15
2 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ АНАЛІЗУ ЕФЕКТИВНОСТІ СИСТЕМ ПОШУКУ РОБОТИ МЕТОДАМИ DATA SCIENCE	16
2.1 Розробка методів формування наборів даних для аналізу ефективності систем пошуку роботи.....	22
2.2 Визначення вимог до моделей отримання оцінок ефективності працевлаштування на ринку праці.....	23
2.3 Розробка моделей формування рекомендацій, отримання прогнозів на ринку праці.....	24
2.4 Висновки за розділом 2.....	32
3 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДЕЛЕЙ ОЦІНКИ ЕФЕКТИВНОСТІ СИСТЕМ ПОШУКУ РОБОТИ	33
3.1 Програмна реалізація оцінки ефективності працевлаштування з урахуванням характеристик претендента, досвіду роботи, наявності документів про освіту, сертифікатів стажування, наявності демонстраційних зразків роботи.....	33
3.2 Експериментальна реалізація методу аналізу моделювання ефективності систем пошуку роботи.....	35

3.3 Експериментальна побудова оцінок ефективності систем пошуку роботи	53
3.4 Висновки за розділом 3.....	57
ВИСНОВКИ	58
ПЕРЕЛІК ДЖЕРЕЛ.....	59
ДОДАТОК А Текст програми	66



ВСТУП

На сьогодні ринок праці характеризується постійними змінами в засадах, попиту і пропозиції, зростає кількість вимог до претендентів. Тому пошук роботи претендентами та заповнення вакансій роботодавцями значно ускладнюється.

Десяток років тому основними засобами для працевлаштування та пошуку кадрів роботодавцями були газети і журнали. Сьогодні, крім газет і журналів, для пошуку роботи використовуються пошукові сервіси загального призначення (інтернет-браузери, соціальні мережі) так і спеціального - спеціалізовані веб-сайти, професійні соціальні мережі [1,2].

Числові оцінки кількості користувачів, які визначають обсяг наборів даних які обробляються в системах пошуку роботи наступні. Сайт www.jobs.ua сьогодні входить в трійку лідерів на українському інтернет-ринку праці, надає комплекс рекламних та інформаційних послуг [3]. Основні користувачі сайту Jobs.ua - це претенденти в пошуку потрібної їм роботи і роботодавці, що підбирають собі персонал. Кожен день вакансії на сайті Jobs.ua переглядає понад 70 000 претендентів. Важливий показник їх зацікавленості сайтом - щохвилине оновлення і доповнення бази вакансій та наступні показники:

- понад 255 527 активних резюме,
- понад 50 300 відвідувачів на добу,
- понад 160 910 переглядів на добу,
- понад 25 384 активних вакансій,
- понад 1 070 000 відвідувачів на місяць,
- понад 3 428 000 переглядів на місяць.

Наведені числові характеристики підтверджують актуальність спеціалізованих додатків пошуку роботи або платформ які дозволяють обрати роботу або кандидатів з урахуванням інтересів і побажань даних користувачів.

Аналіз гендерної структури користувачів дозволяє зробити висновок, що ІТ-продукти пошуку роботи актуальні в рівній мірі як для чоловіків так і жінок (рис.1.1).

Аналіз структури користувачів за географією України дозволяє зробити висновок, що ІТ-продукти пошуку роботи найбільше актуальні у користувачів Києва та Київської області в східних та деяких західних областях (рис.1.2).

Аналіз структури користувачів за рівнем освіти дозволяє зробити висновок, що ІТ-продукти пошуку роботи актуальні для двох основних підгрупи цільової аудиторії: у користувачів з середньою (38,1%) та вищою освітою (40,4%) (рис.1.3).

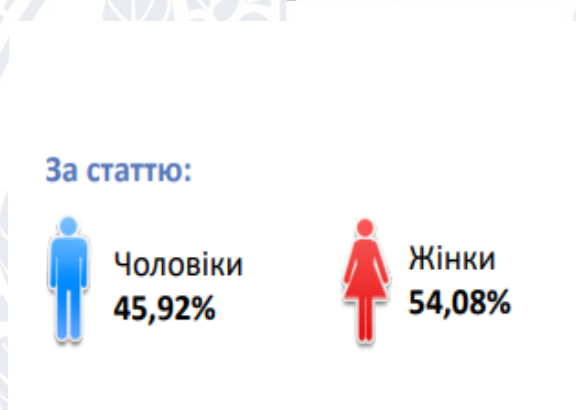


Рис.1.1 Гендерна структура користувачів додатку пошуку роботи www.jobs.ua



Рис.1.2 Структура користувачів додатку пошуку роботи за географією www.jobs.ua



Рис.1.3 Структура користувачів додатку пошуку роботи за рівнем освіти www.jobs.ua

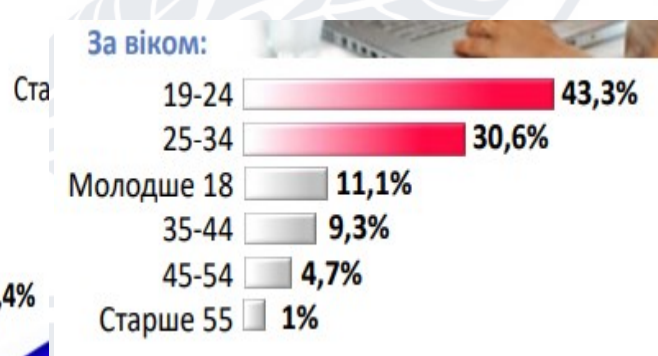


Рис.1.4 Вікова структура користувачів додатку пошуку роботи www.jobs.ua

Аналіз структури користувачів за віком дозволяє зробити висновок, що ІТ-продукти пошуку роботи актуальні для двох основних підгрупи цільової аудиторії: у користувачів у віці 25-34 років (30,6%) та у віці 19-24 років (43,3%) (рис.1.4).

Популярність та затребуваність додатків, сервісів, сайтів пошуку роботи зростають на фоні постійного збільшення потреби у працевлаштуванні населення у більшості країн світу, а ситуація з карантинними заходами, запровадженими у зв'язку з пандемією COVID-19 спричинила стрімкий ріст рівня безробіття, що призвело до попиту на сервіси пошуку роботи та високих темпів зростання кількості користувачів таких сервісів.

Числові оцінки кількості показників, атрибутів пошуку, які визначають другий вимір набору даних складають декілька десятків.

1. Характеристики оголошення пошуку роботи: регіон, тип зайнятості, галузь/вид діяльності, компанія, мова, заробітна плата очікувана, стаж роботи, рівень посади, тип роботи (он-лайн/ оф-лайн), тип зайнятості (повний/частковий), вік претендента, режим роботи.

2. Технічні характеристики оголошення про пошук роботи: термінове оголошення, гаряча пропозиція, кольоровий блок в оголошенні, VSP-компанія, автопоновлення оголошення, платні публікації, банерна реклама в оголошенні, дата створення оголошення.

3. Характеристики процесу пошуку роботи на сайті: дата публікації оголошення роботодавцем, заробітна плата середня, заробітна плата мінімальна, заробітна плата медіанна, вік претендента мінімальний, вік претендента максимальний, вік претендента медіанний.

4. Характеристики вакансій: регіон, тип зайнятості, галузь/вид діяльності, компанія, мова, заробітна плата мінімальна, заробітна плата максимальна, стаж роботи, рівень посади, тип роботи (он-лайн/ оф-лайн), вік претендента, режим роботи, дата публікації оголошення роботодавцем.

5. Характеристики компаній, підприємств, організацій, які пропонують вакансії: тип компанії за видом власності, галузь, рейтинг компанії, розмір компанії, рік створення, локація, назва компанії.

6. Характеристики умов праці: тип зайнятості (повний/частковий); тип роботи (он-лайн/ оф-лайн); режим роботи.

7. Характеристики вакансій за результатами роботи системи за період: закрита/ не закрита вакансія за певний період (тиждень/місяць); час від появи оголошення до закриття вакансії; кількість заяв на вакансію від претендентів за певний період (тиждень/місяць).

Ефективність пошуку роботи можна оцінювати для вакансій роботодавців та резюме в цілому по всій системі пошуку, по регіонам, по спеціальностям, по віку. При великій кількості параметрів пошуку виникає необхідність надання рекомендацій (ранжування пропозицій отриманих за автоматичним запитом) з урахуванням особливостей роботодавця або претендента на роботу.

На сьогодні вже існує багато готових інструментів, які надають можливість збирати, аналізувати дані та приводити їх до потрібного вигляду для подальшого опрацювання та/або збереження. Проте в більшості вони реалізують автоматичні запити користувачів та не надають рекомендацій на основі аналізу ефективності пошуку попередніми користувачами.

Таким чином на сьогодні актуальною є задача дослідження ефективності таких інформаційно-пошукових систем, як веб-сайти пошуку роботи, а також розробка моделей, методів, зокрема інтелектуалізації функцій даних сайтів з метою підвищення ефективності в частині персоналізованих рекомендацій для претендента.

Об'єктом дослідження даної роботи є процес оцінювання ефективності систем пошуку роботи.

Предметом дослідження даної роботи є методи Data Science для аналізу статистичних даних, які можуть використовуватися для оцінювання ефективності систем пошуку роботи.

Мета роботи полягає в дослідженні методів оцінки ефективності систем пошуку роботи України за допомогою методів Data Science та розробка моделей для формування рекомендацій для роботодавця та претендентів на роботу.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- проаналізувати сучасні інформаційні та рекомендаційні системи пошуку роботи;
- дослідити набори даних систем пошуку роботи України;
- дослідити методи обробки та аналізу даних систем пошуку роботи;
- дослідження моделей для формування рекомендацій, отримання прогнозів на ринку праці;
- обґрунтувати вибір методів аналізу даних для визначення ефективності систем пошуку роботи;
- побудувати моделі оцінки ефективності системи пошуку роботи України.

Методи дослідження: системний підхід для при формалізації та класифікації факторів системи пошуку роботи; методи регресійного аналізу, кластеризації та класифікації (з використанням бустингу) та методи статистичного та кореляційного аналізу великих обсягів даних за набором показників і кількістю спостережень з використанням сучасних програмних засобів (R, Python) та технологій (Google Collab).

Практичне значення результатів роботи полягає в тому, що розроблена концептуальна схема моделювання та обрані методи багатовимірного аналізу даних дозволяють проводити розрахунок і оцінки ефективності систем пошуку роботи.

Основні результати за темою магістерської роботи пройшли апробацію:

- на всеукраїнській науково-практичній конференції «Комп'ютерні технології обробки даних» Донецький національний університет імені Василя Стуса, 2021 р., 2022р [4, 5];

РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Дослідження існуючих систем пошуку роботи за цільовою аудиторією, функціоналом, характеристиками наборів даних про вакансії і претендентів

Для дослідження існуючих інформаційних та рекомендаційних систем пошуку роботи проведемо порівняльний аналіз за видом програмного продукту; функцій (основних та додаткових), за цільовою аудиторією, спеціалізацією, за даними які обробляються в системі (показників пошуку, статистика роботи сайту), переваг і недоліків.

На сьогодні існують різні інформаційні ресурси: веб-сайти, соціальні мережи що допомагають у здійсненні працевлаштування. Основні з них:

- work.ua
- jobs.ua
- dou.ua
- ua.endeed.com
- LinkedIn
- djinni.co
- Monster.com
- Справді
- USAJobs

Загальна функція даних систем пошуку роботи: надання доступу до бази з вакансіями та резюме і оснащення необхідними інструментами для роботи з цією базою.

Розглянемо найбільші системи пошуку, які є лідерами за кількістю користувачів.

Так, наприклад, сайт пошуку роботи «work.ua» – найбільший в Україні сайт для пошуку співробітників. Дана платформа має розширену систему пошуку вакансії за категоріями, містами, компаніями, посадами, яка дозволяє

швидко отримувати результати, що максимально відповідають введеному запиту (див. рис. 1.5). На сайті запропонована зручна форма додавання резюме, яка дозволяє кожному, хто шукає роботу, вірно презентувати себе роботодавцю.

Є можливість перегляду новин та статей, що стосуються сфери працевлаштування. Якщо користувач перейде до вкладки «Роботодавцю», він матиме можливість переглянути наявні резюме за категоріями або містами. За необхідності може розмістити власну вакансію. Щодо зручності цієї платформи для студента, то в пошуку вакансії користувач може натиснути на посилання «Робота для студента», щоб побачити вакансії, що створені для пошуку працівника-студента, проте у більшості вакансій вказано, що необхідний досвід роботи для працівника від 1 року. Спеціальних модулів сайту для пошуку робочого місця студентами без досвіду роботи – немає.

У той же час, багато людей щомісяця, щоб знайти роботу відвідують вебсайт «rabota.ua», а провідні компанії України розміщують свої вакансії і знаходять співробітників.

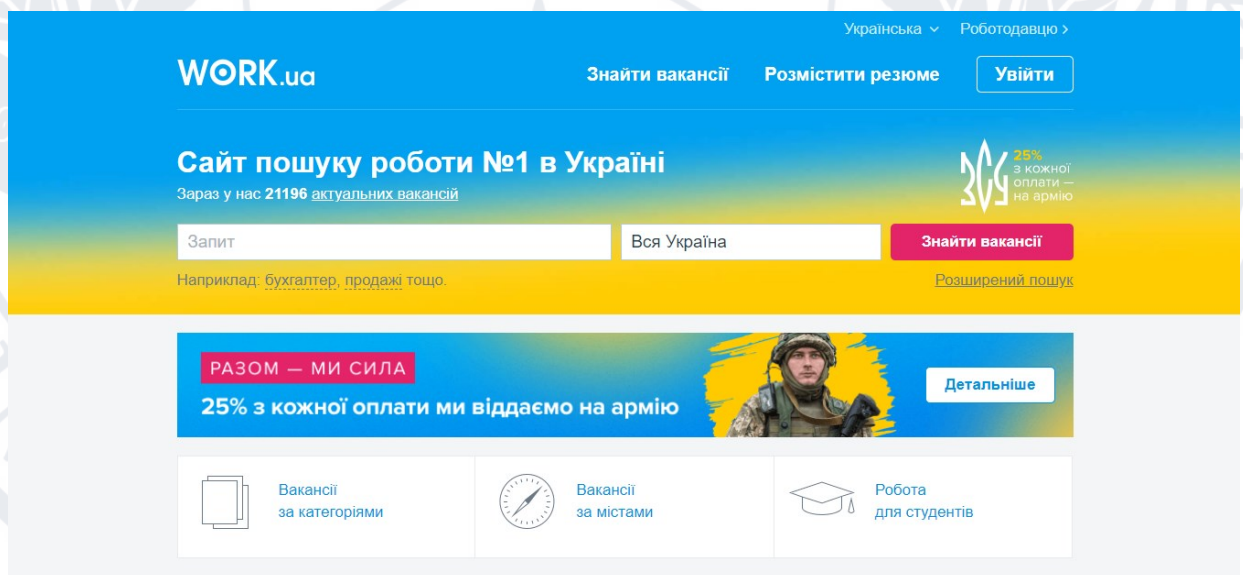


Рис. 1.5 – Головна сторінка сайту work.ua

Rabota.ua - перший український сайт з пошуку персоналу, який 22 роки тому почав працювати на інформаційному просторі українського інтернету (див. рис. 1.6). 75% українських інтернет-користувачів вибирають саме цей

майданчик для пошуку роботи. На даному сайті доступні всі стандартні функції:

- заповнення анкети,
- сортування вакансій по категоріях,
- даті розміщення;
- вибору регіону.

Для спрощення процесу розробники передбачили наявність логотипів відомих компаній на головній сторінці. Завдяки цьому вакансію саме на потрібну посаду буде знайдено ще швидше.

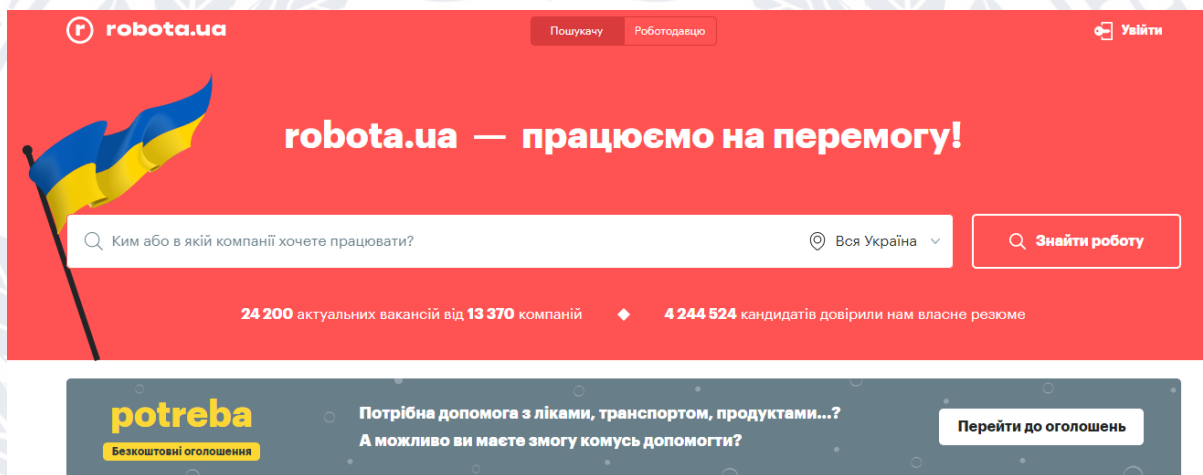


Рис. 1.6 – Головна сторінка сайту rabota.ua

Для пошуку роботи студентами реалізований модуль «Робота для студентів», який надає можливість перегляду попередньо відібраних вакансій, а також пошуку з фільтрами по цих вакансіях.

Вищезгаданий веб-сайт має аналітичні функції і надає наступну інформацію:

- статистику по Україні про розподіл зарплати (за вакансіями та резюме);
- про середню заробітну плату;
- про кількість вакансій та активних кандидатів;

- про пропозиції компаній окрім зарплати, такі як бонуси, розташування офісу, знижки для працівників тощо; про топові сфери та професії ринку праці.

На сайті пошуку роботи Джині кандидат описує свій досвід, очікування від роботи та побажання по зарплаті, а компанії пропонують вакансії. Кандидат вирішує, кому і коли відкрити контакти.

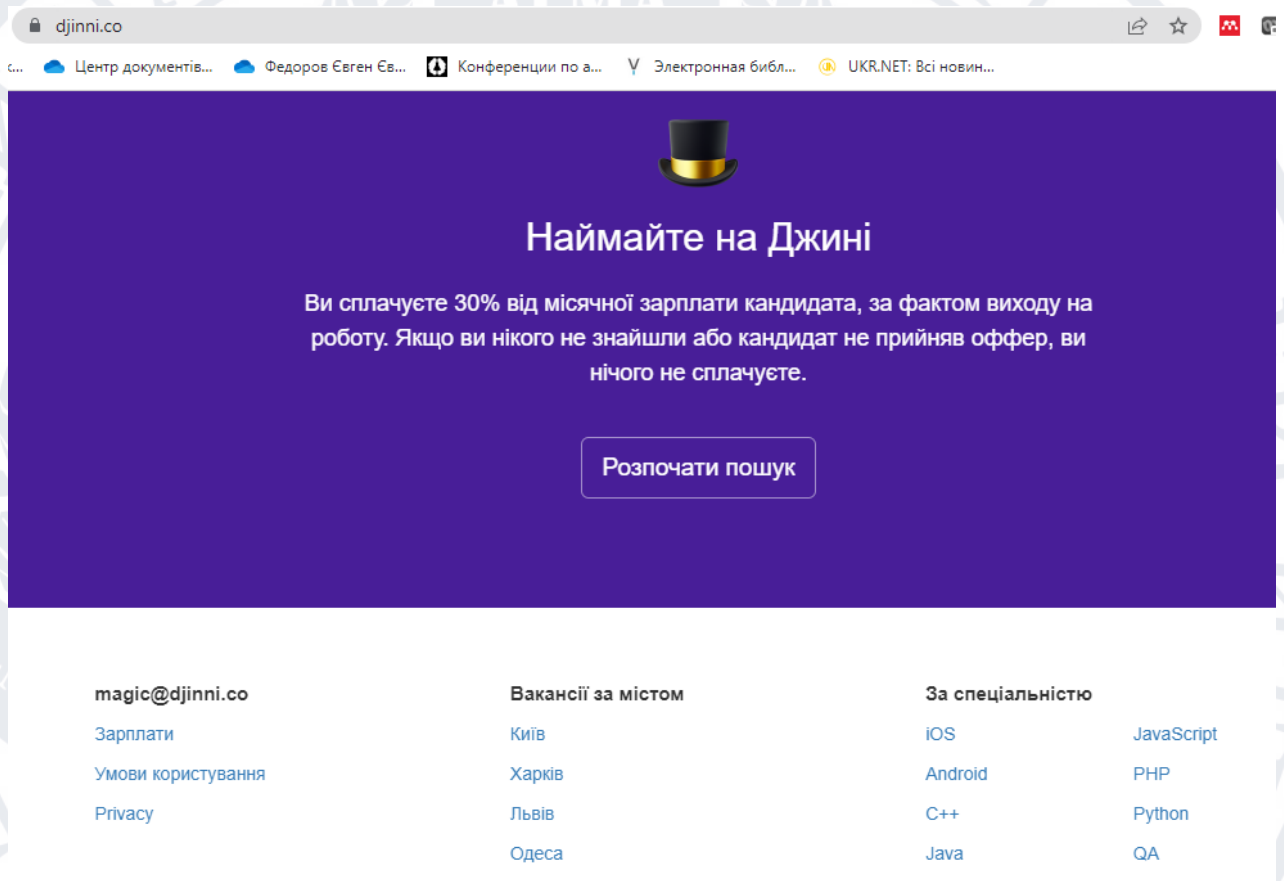


Рис. 1.7 – Головна сторінка сайту Джині

Monster.com – одна з найстаріших пошукових систем роботи в Інтернеті. Хоча деяка його корисність зменшилася за ці роки через відсутність належної фільтрації та занадто велику кількість публікацій спам-рекрутерів, це все ще важливий сайт, на якому можна шукати роботу. Є функції звуження пошуку роботи Monster:

- за місцем розташування,
- ключовими словами;

– роботодавцем.

Існують додаткові функції для пошуку роботи, таких як мережеві дошки, попередження про пошук роботи та розміщення резюме в Інтернеті.

Роботодавці також можуть використовувати Monster.com для пошуку співробітників з метою:

- розширити свій перелік посад з найму,
- знайти нового штатного працівника;
- контрактного працівника;
- зібрати пул потенційних претендентів на майбутню посаду.

Переваги. Наявність наступних функцій

- Підписка на попередження про роботу.
- Пошук роботи в будь-якому місці.
- Хороша бачимість у рекрутерів.
- Дозвіл на розміщення резюме в мережі.

Недоліки

- Багато спам-повідомлень від тіньових рекрутерів.
- Загалом слабка курація сайту та його змісту.

Indeed.com – це механізм пошуку роботи з можливістю скласти резюме та подати його на місце для пошуку роботодавцями ключових слів, робочих місць, ніш тощо. Інші функції включають можливість фільтрування робочих місць за оцінкою заробітної плати та можливість виконувати високотехнологічний пошук.

Цей веб-сайт розкриває широкий спектр робочих місць та полів, яких зазвичай не має у більшості подібних пошукових систем, і вони добре роблять свою функцію пошуку якомога простішою у використанні. Наприклад, ви можете шукати стажування та тимчасові роботи, або контракт, штатний розклад тощо. Ці розширені можливості пошуку реалізовані на сторінці розширеного пошуку “Справді”. Тобто дана система має перевагу порівняно з іншими системами пошуку роботи за переліком показників наборів даних про вакансії та претендентів, які обробляються в системі.

Також наявні функції які дозволяють користувачу підлаштувати систему під свої потреби. Максимально просто претенденту відстежувати роботу; все, що потрібно зробити, це створити обліковий запис, і кожна робота, на яку претендент подав заявку через Indeed.com або до якої виявив зацікавленість, буде збережена у профілі користувача. Щоб оперативно отримувати інформацію про можливості роботи, наявна функція підписки на сповіщення про роботу за допомогою електронної пошти (щоденні або щотижневі сповіщення) і налаштування їх для певних ключових слів та конкретного місяця.

Переваги

- Глибина в нішевих галузях;
- Бонусні функції, якщо потрібно скласти резюме з нуля.
- Хороший вплив рекрутерів по всій країні.

Недоліки:

- Досить консервативний дизайн та функціональність.
- Сповіщення електронною поштою можуть бути не релевантними.

CareerBuilder має наступні функції для претендентів:

- можливість знайти роботу,
- опублікувати резюме,
- створювати попередження про роботу,
- отримувати поради щодо роботи та ресурси,
- шукати ярмарки вакансій та інш.

Мільйони унікальних відвідувачів використовують CareerBuilder для пошуку нової роботи та отримання кар'єрних порад. Це система пошуку роботи, яка пропонує багато ресурсів для тих, хто шукає роботу; зокрема список спільнот з пошуку роботи. Після первинного пошуку роботи є можливість відфільтрувати результати за типом роботи, датою опублікування, оплатою та підтримкою функції «Легке подання заявки».

Переваги:

- велика громада з гідною міжнародною експозицією;

- додаткові ресурси, які допоможуть шукачам роботи вчитися та знаходити ідеї кар'єри;

- підтримка завантаження резюме.

Недоліки:

- модель “побудувати кар'єру” не дуже оптимізована для людей, які просто хочуть знайти роботу;

- не завжди найкращий сайт для рекрутерів з високими можливостями.

SimplyHired також пропонує дуже унікальний досвід пошуку роботи: користувач навчає сайт роботи, оцінюючи робочі місця, які його цікавлять.

Тобто функція пошуку роботи реалізована на основі машинного навчання, тобто є інтелектуальною, хоча і в дуже простій версії. SimplyHired також має аналітичні функції, які дозволяють досліджувати зарплати, додавати вакансії на карту вакансій, створювати своє власне резюме та переглядати досить докладні профілі різних компаній.

Дана система пошуку роботи фокусується на місцевих переліках вакансій. Ви можете переглядати місто, поштовий індекс або штат, щоб знайти потрібну вам роботу. Є унікальна функція на цьому веб-сайті, яка полягає у фільтруванні списку вакансій, щоб знайти лише комісійні. Також є типові фільтри типу роботи, такі як пошук неповного робочого дня, контрактна робота, штатна робота тощо.

Переваги

- інноваційний алгоритм швидкості роботи, який відображає можливості, які вам сподобаються;

- хороша локалізація.

Недоліки:

- не є прем'єрним місцем призначення для рекрутерів вищого рівня;

- алгоритм навчання може не виявити деякі потенційно цікаві можливості.

LinkedIn.com поєднує в собі найкраще з двох типів пошукових систем: можливість шукати роботу в Інтернеті за допомогою своєї системи пошуку роботи та можливість спілкуватися з друзями-однодумцями та приватними особами, щоб поглибити пошук вашої роботи. Оголошення про роботу LinkedIn мають найвищу якість, і якщо ви пов'язані з кимось, хто вже знає про цю конкретну роботу, вам вдасться вступити ще до того, як подати резюме. Тільки обов'язково створіть чудовий профіль LinkedIn. Для пошуку роботи в LinkedIn, є використовувати широкий спектр фільтрів. Сюди входять дата опублікування, проста заявка, тип роботи, місцезнаходження, компанія, галузь, функція роботи, списки, які мають менше 10 претендентів, поїздки на роботу (наприклад, віддалена робота), звання, переваги та рівень досвіду.

Переваги:

- міцні зв'язки з професійною мережею;
- додатки в один клік;
- деякі переліки показують, скільки інших заявників зараз;
- велика кількість варіантів фільтрації.

Недоліки:

Відкриття вашого профілю рекрутерами може відкрити ваш профіль рекрутерам на вашій поточній роботі.

Відмінною особливістю сайту для пошуку роботи Glassdoor є його мета – підвищити прозорість робочого місця. Окрім мільйонів списків вакансій – це огляди компаній, рейтинги схвалення директора, огляди співбесід та питання, а також інші відомості про компанії, які пропонують роботу. Glassdoor – це міжнародна система пошуку роботи, тому база даних системи дозволяє знайти роботу в США, а також Канаді, Мексиці, Бразилії, Франції, Австралії та інших місцях. Два особливих варіанти фільтрації, на які варто виокремити, – це початковий рівень та учень / стажер; більшість пошукових систем не пропонують їх. Реалізований фільтр переліку робочих місць за рейтингом компанії, галуззю, функціями роботи, розміром компанії та іншими. Також доступні типові варіанти фільтрування, такі як дата публікації, зарплата та

відстань. База даних Glassdoor також містить показник можливості віддаленої роботи, який розширює можливості пошуку роботи в Інтернеті. Просто виберіть WFH або віддалений з розділу фільтрації.

Переваги:

- реалізація аналітичної функції, яка надає можливість отримання рейтингів компаній;
- розширена база даних по місту роботу, яка надає можливості пошуку роботи по всьому світу.
- реалізовано багато варіантів фільтрації;
- зворотній зв'язок, у формі відгуків про компанії;
- детальні попередження про роботу / компанію.

Недоліки:

Багато додаткових послуг, які можуть бути зайвими.

Захоплення – одна з систем пошуку роботи для підлітків. Ця пошукова система для вакансій має показник «Підліток», який дозволяє відфільтрувати вакансії на які приймають молодих людей. Тобто, не потрібно вводити пошуковий термін на цьому сайті; просто достатньо орати фільтр для підлітків та будь-який інший, наприклад, відстань чи галузь, та отримаєте всі завдання для підлітків. Звичайно, сайт знаходить роботу і для інших віків. Існує фільтр для термінового найму – один для додатків у 1 клік, а інший – для сезонної роботи, роботи за сумісництвом або повного робочого дня. Усі завдання у списку також можна відсортувати за тими, що були нещодавно додані, щоб збільшити шанси бути першим заявником та отримати посаду.

Переваги

- чіткий і простий дизайн веб-сайту.
- унікальні варіанти фільтрації.

Недоліки:

- відсутні функції на деяких сайтах вакансій.
- потрібен обліковий запис користувача, який застосовується до вакансій.

База даних Craigslist містить всілякі цікаві роботи. Потрібно знайти своє місто, а потім в розділ переліку місцевих робочих місць. Тут існує безліч категорій робочих місць, включаючи освіту, маркетинг, уряд, транспорт, медіа, адміністратора, офіс, технічну підтримку тощо. Також є можливість встановити різні канали RSS, які стосуються будь-якої роботи, яку ви могли б шукати, у будь-якому місці. Робота в Craigslist може бути шахрайством, тому потрібно бути обережним та використовувати здоровий глузд, відповідаючи на списки вакансій.

Переваги:

- широка база даних для пошуку місцевих дивних робіт.

Недоліки:

- відсутність перевірки подань.
- похмурий дизайн веб-сайту.

CarabilityJOBS – це сайт роботи для людей з обмеженими можливостями. Насправді він стверджує, що є лише система пошуку роботи, де 100 відсотків списків – від роботодавців, які активно шукають лише людей з обмеженими можливостями. Як і більшість цих найпопулярніших сайтів із вакансіями, sposobnostJOBS може сортувати список вакансій за посадою, компанією, місцем розташування та датою розміщення. Також є можливість відфільтрувати список вакансій за функцією, галуззю, типом роботи, штатом та милями (до 100) з будь-якого поштового індексу. Це ще одна система пошуку роботи з унікальними фільтрами типу роботи. Можна вибрати тимчасово найнятого або добровольця, щоб знайти роботу, позначену як таку, на додаток до звичайних фільтрів для повного / неповного робочого часу, роботи за контрактом тощо.

Переваги:

- спеціалізований для людей з обмеженими можливостями;
- роботодавці можуть вас знайти, якщо ви опублікуєте своє резюме;
- підтримує попередження про роботу.

Недоліки:

Потрібно створити обліковий запис користувача перед переглядом завдань.

1.2 Дослідження реалізації зворотнього зв'язку щодо ефективності працевлаштування. Методи оцінки ефективності систем пошуку роботи

Щодо отримання зворотнього зв'язку можна виокремити два основних підходи: опитування (роботодавців, претендентів на роботу) і статистика сайтів.

Опитування включає аналіз структури цільової аудиторії користувачів за наступними характеристиками: сфера діяльності, посада, регіон, довід роботи і може виконуватись за різними критеріями.

Опитування за джерелом успішного пошуку.

За результатами пошуку: по посаді, по регіону, за досвідом.

Опитування з метою виявлення поведінки користувачів: наприклад, чи моніторять вони сайти пошуку роботи і з якою метою.

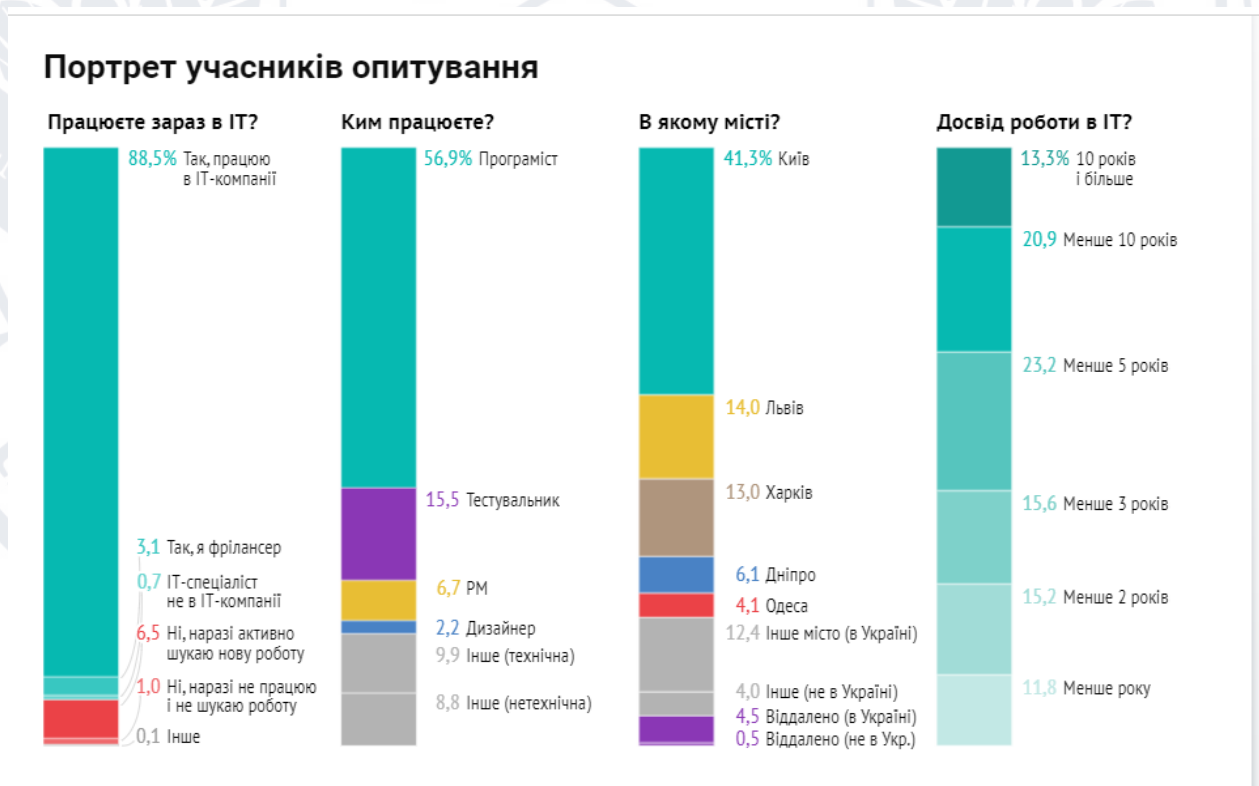


Рис. 1.8 Структура учасників опитування щодо працевлаштування

Як ви знайшли своє поточне місце роботи?

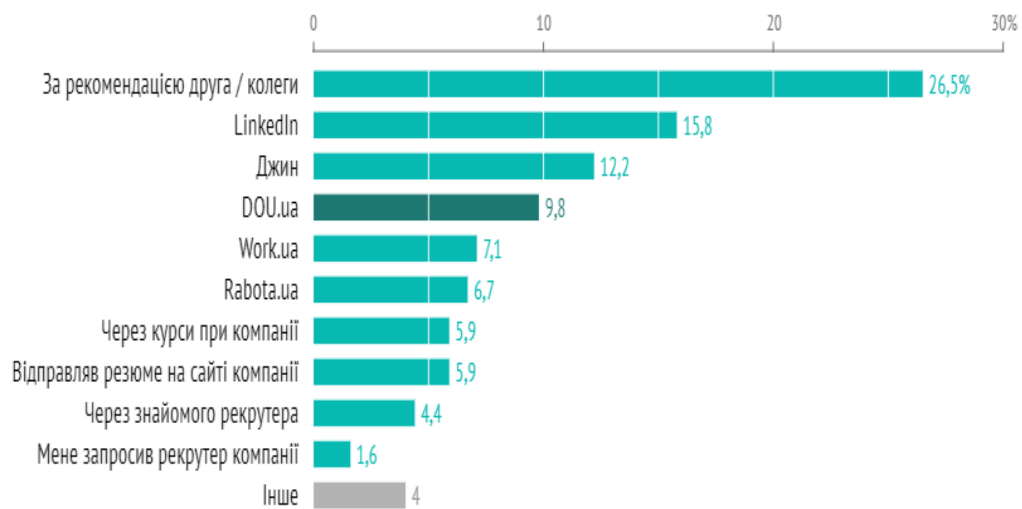


Рис. 1.9 Структура відповідей по результативних системах пошуку роботи

Як ви знайшли своє поточне місце роботи: за посадами

	Програміст	Тестувальник РМ	Дизайнер	Інше (технічна)	Інше (нетехнічна)
За рекомендацією друга / колеги	25,3	26,2	31,1	24,7	29,5
LinkedIn	18,3	13,8	10,9	10,1	14,1
Джин	14,0	12,5	10,5	11,2	6,5
DOU.ua	9,3	10,8	12,4	7,9	8,1
Через курси при компанії	6,4	11,2	1,1	3,4	1,1
Відправляв резюме на сайті компанії	5,9	6,7	6,0	5,6	4,8
Rabota.ua	5,4	6,1	7,1	10,1	10,6
Work.ua	5,2	5,9	9,7	15,7	11,1
Через знайомого рекрутера	4,6	3,4	7,5	1,1	3,0
Мене запросив рекрутер компанії	1,6	1,0	1,5	4,5	2,0
Інше	4,0	2,4	2,2	5,6	7,2

Рис. 1.10 Структура відповідей по результативних системах пошуку роботи по посадах у сфері ІТ

Як ви знайшли своє поточне місце роботи: за містами

	% Київ	Дніпро	Львів	Одеса	Харків	Інше місто (Укр.)	Інше місто (не Укр.)	Віддалено (Укр.)
За рекомендацією друга / колеги	24,1	28,6	24,6	29,5	24,0	38,0	23,0	26,7
Джин	16,2	9,4	8,4	13,9	13,4	5,5	1,2	13,3
LinkedIn	15,6	10,2	20,5	11,4	15,4	8,7	38,5	15,0
DOU.ua	12,0	11,0	8,5	9,6	9,8	5,3	1,9	13,3
Rabota.ua	9,4	2,9	3,9	5,4	7,1	5,3	1,9	5,0
Work.ua	6,0	13,9	3,4	8,4	7,7	11,3	–	11,1
Відправляв резюме на сайті компанії	5,3	3,7	6,6	4,8	6,7	6,3	16,1	0,6
Через знайомого рекрутера	3,4	2,9	6,9	6,0	7,3	3,4	3,7	1,7
Через курси при компанії	3,0	12,2	11,6	6,6	5,6	10,3	–	1,1
Мене запросив рекрутер компанії	1,9	2,0	0,9	1,8	1,2	1,8	1,2	1,1
Upwork.com	0,2	–	0,2	–	0,2	0,8	–	6,7
Інше	2,9	3,2	4,6	2,4	1,7	3,4	12,4	4,5

Рис. 1.11 Структура відповідей по результативних системах пошуку роботи по містах України

Як ви знайшли своє поточне місце роботи: за досвідом

	% Менше року	<2 років	<3 років	<5 років	<10 років	10+ років
За рекомендацією друга / колеги	22,5	23,7	24,7	28,0	28,4	29,8
Через курси при компанії	16,6	9,8	6,7	2,8	2,9	1,5
Work.ua	11,7	10,1	7,7	6,1	4,9	3,9
Rabota.ua	10,6	10,6	7,9	4,7	4,3	4,5
DOU.ua	10,0	13,6	9,3	9,9	8,0	8,8
Джин	8,1	12,3	14,3	13,8	12,4	10,3
Відправляв резюме на сайті компанії	6,2	4,3	5,8	6,0	7,0	6,0
LinkedIn	3,6	9,8	16,8	18,7	20,0	20,3
Через знайомого рекрутера	1,9	2,0	3,8	4,5	6,6	6,6
Мене запросив рекрутер компанії	0,6	0,8	0,6	2,2	2,3	2,3
Upwork.com	–	–	0,3	0,6	0,6	1,7
Інше	8,1	3,0	2,0	2,6	2,7	4,4

Рис. 1.12 Структура відповідей по результативних системах пошуку роботи по стажу роботи

Чи моніторите ви ринок праці, незалежно від наявності постійного місця роботи?

[Посада](#) [Місто](#) [Досвід](#) [Статус](#)

	%	Програміст	Тестувальник РМ	Дизайнер	Інше (технічна)	Інше (нетехнічна)
Читаю зарплатні опитування DOU	78,6	84,6	71,9	66,3	70,3	63,2
Переглядаю вакансії на DOU	54,7	61,3	57,7	62,9	53,9	60,9
Відповідаю рекрутерам на LinkedIn	51,3	52,6	51,7	39,3	49,6	39,7
Читаю зарплатну розсилку Джина	39,2	42,5	32,2	42,7	28,0	22,7
Переглядаю вакансії на інших сайтах	23,6	24,1	25,5	37,1	30,5	39,9
Ходжу на співбесіди, щоби бути у формі	15,8	15,1	15,7	10,1	14,9	13,9
Відгукуюся на вакансії	15,3	14,6	15,7	23,6	17,6	19,5
Включаю профіль на Джині	11,4	14,3	18,7	22,5	10,8	12,7
Ходжу на співбесіди з метою надбавки	5,4	4,7	1,5	1,1	2,8	1,7
Не роблю нічого з вищезазначеного	4,7	3,0	7,9	9,0	6,8	6,5

Рис.1.13 Результати опитування за видами моніторингу ринку праці

Отже, в наведеному прикладі, щодо каналів пошуку роботи в ІТ-сфері, то найрезультативнішими виявились особисті рекомендації: 26% опитаних відповіли, що знайшли поточне місце роботи за рекомендацією друга або колеги; а ще 4% запросив знайомий рекрутер. Наступним за результативністю є LinkedIn: за допомогою цієї мережі працевлаштувались 16% опитаних.

Профільні сайти також активно сприяють у пошуках роботи в ІТ: на Джині та DOU знайшли роботу 12% та 10% опитаних відповідно. Дещо менш ефективним способом пошуку роботи в ІТ виявились сайти пошуку роботи Work.ua (7%) та Rabota.ua (7%).

Курси та навчальні центри при компанії надали роботу 6% опитаних. Ще для 6% успішною виявилася спроба надіслати резюме на сайт компанії. А 2% відповіли, що з ними зв'язався незнайомий рекрутер із запрошенням на роботу через скайп, особисту пошту або за допомогою телефонного дзвінка.

Основним інструментом для пошуку роботи та відбору претендентів на роботу є спеціалізовані соціальні мережі та веб-сайти.

Який з каналів пошуку роботи вам здається найбільш ефективним

Посада Місто Досвід Статус Планує цього року змінити роботу

	%	Програміст	QA	Analyst	PM	Дизайнер	DevOps	Інше (технічна)	Інше (нетехнічна)
Djinni.co	73	75	71	65	63	66	60	61	
LinkedIn	66	55	62	61	50	66	53	48	
Через знайомих	46	44	46	51	46	44	47	46	
DOU.ua	37	46	48	44	38	38	44	49	
Розсилка резюме в компанії	24	28	32	26	32	22	33	34	
Work.ua	6	9	12	10	16	4	16	17	
Robota.ua	4	8	13	9	13	5	13	15	
Skyworker	0	0	–	–	–	–	–	0	
Upwork	0	0	–	1	1	1	0	–	
Соцмережі	0	0	–	–	–	1	1	1	

Рис.1.14 Результати опитування щодо рівня ефективності каналів пошуку роботи серед IT-фахівців

За результатами наведеного опитування користувачі систем пошуку роботи вважають найефективнішими для пошуку роботи Djinni й LinkedIn

Оцінка різних каналів для пошуку роботи в українському IT практично не змінилася за рік — Djinni та LinkedIn вважають найефективнішими.

Djinni стали дещо частіше оцінювати як ефективний канал (70% проти 68% рік тому), а LinkedIn — трохи рідше (60% проти 63% рік тому), що вивело Djinni на перше місце за цим параметром.

Дещо частіше Djinni називають ефективним каналом для пошуку роботи розробники, QA, спеціалісти з Data Science, а також HR'и. LinkedIn популярний серед усіх спеціалізацій, хоча HR-фахівці трохи рідше за інших вважають його ефективним. DOU частіше за інших називали ефективним HR'и, аналітики і спеціалісти з Data Science та тестувальники. Інші українські сайти з пошуку роботи (Work.ua, Robota.ua) частіше вважають ефективними дизайнери й аналітики.

Вище за інших визначають ефективність Djinni та LinkedIn спеціалісти з досвідом від 1 року. Менш досвідчені фахівці дещо краще за інших оцінюють DOU, інші українські сайти з пошуку роботи, а також розсилку резюме безпосередньо в компанії.

Другий канал інформації щодо отримання зворотнього зв'язку щодо ефективності систем пошуку роботи це офіційні результати працевлаштування користувачів даних систем.

За результатами опитування або статистики розробляються функції, які підвищують ефективність вакансій. Наприклад, сайт Jobs.ua пропонує роботодавцям скористатись спеціальними послугами, що підвищують відсоток відгуку до 500%: термінова, гаряча, кольоровий блок, VSP-компанія, автопоновлення, платні публікації, банерна реклама. Автооновлення вакансій.

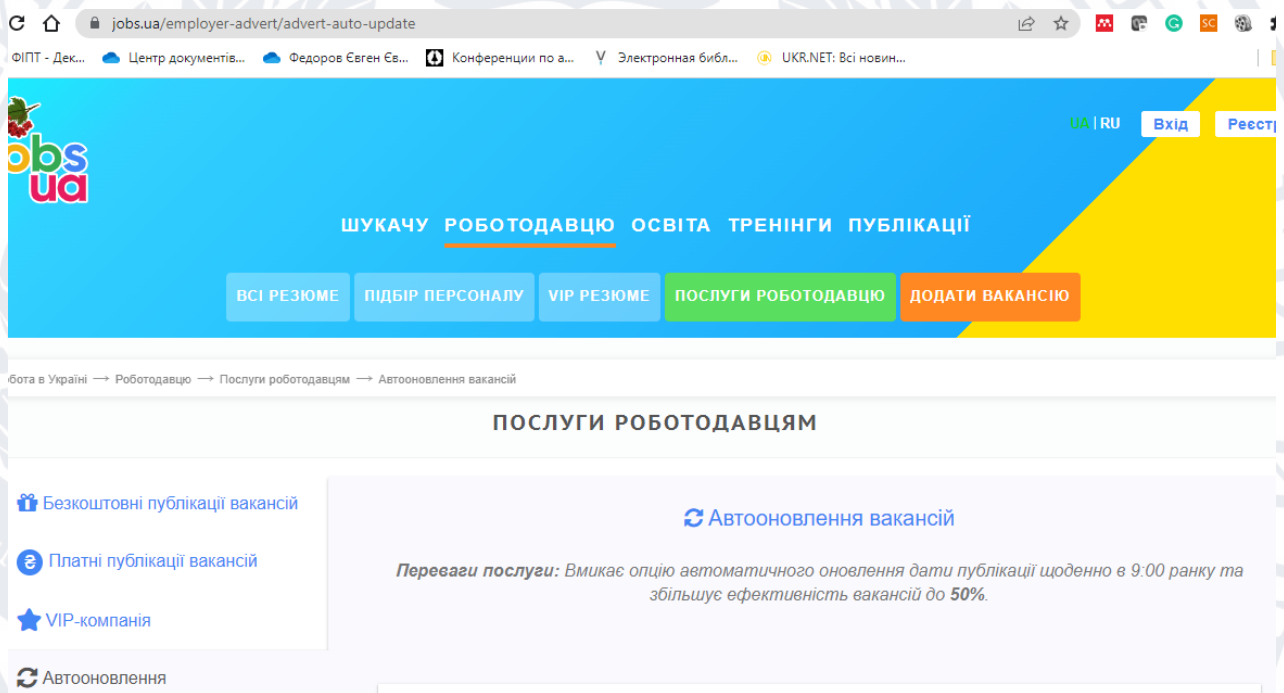


Рис.1.15 Приклад реалізації функції спрямованої на підвищення ефективності оголошень про вакансію

В роботі [17] розроблено опитування, яке зосереджується на поведінці людей у пошуку роботи незалежно від їх статусу робочої сили, і проводиться його щорічно, починаючи з 2013 року. Опитування використовується, щоб вивчити зв'язок між пошуковими зусиллями та результатами для зайнятих і непрацюючих. Виокремлені наступні три важливі факти: (1) пошук на робочому місці є поширеним і інтенсивнішим на нижчих щаблях службової драбини; (2) зайняті принаймні втричі ефективніші за безробітних у пошуку роботи; і (3) зайняті отримують кращі пропозиції роботи, ніж безробітні.

Розроблена загальна модель рівноваги пошуку на робочому місці з ендогенними пошуковими зусиллями, відкалібрована відповідно до нових фактів і виявлено, що пошукові зусилля найнятих є дуже еластичними.

1.3. Дослідження методів обробки даних в системах пошуку роботи **Алгоритм прямого пошуку**

Це метод прямого перебору всіх сторінок (документів), що є в базі даних пошукової системи. Цей метод дозволяє зі 100 надійністю знайти необхідну інформацію, але він є трудомістким і тому, при обробці великих обсягів даних, пошук буде займати багато часу.

Алгоритм зворотного пошуку (інвертованих індексів)

Для ефективного пошуку у великих обсягах даних всі потужні пошукові системи використовують алгоритм зворотних (інвертованих) індексів.

За цим алгоритмом документи в пошуковій системі перетворюються в текстові файли, що містять перелік всіх слів наявних у документі. Слова в таких списках (індекс-файлах) розташовуються в алфавітному порядку і кожному слову у відповідність призначаються координати його знаходження в документі та параметри, що визначають його статус в документі.

Структура таких індекс-файлів є схожою до алфавітного покажчика слів в технічних або наукових книгах, де наводиться список термінів із зазначенням номерів сторінок, де вони зустрічаються.

Для створення сторінки результатів пошуку дані системи шукають інформацію саме в зворотних індексах документів, які обробляють. Прямі індекси (оригінальний текст документів) системи пошуку теж використовують, наприклад для складання фрагментів опису знайденого документу.

Для пошуку по зворотних індексах документів у базі даних пошукових систем застосовується алгоритм, що дозволяє зменшити обчислювальну складність процесу виявлення потрібних документів (за введеним користувачем пошукового запиту) і процес визначення релевантності

документів знайдених відповідно до запиту. Чим краще документ відповідає запиту, тим вище він буде розташований в пошуковій видачі.

Основним завданням математичної моделі покладеної в основу алгоритму будь-якої пошукової системи є пошук документів (сторінок) у своїй базі зворотних індексів відповідних до даного пошукового запиту і сортування цих знайдених документів у порядку зменшення їх релевантності до пошукового запиту. Використання простої логічної математичної моделі, яка знаходить документ, якщо в ньому зустрічається шукана фраза, не підходить, в силу великої кількості таких документів.

Математична модель, яку використовується пошукові системи, відноситься до класу векторних математичних моделей. В ній використовується поняття ваги документа по відношенню до заданого користувачем запиту.

В базовій векторній моделі вага документа за заданим пошуковим запитом обчислюється за двома основними параметрами: частотою, з якою зустрічається дане слово в аналізованому документі (TF - term frequency) і частотою, наскільки рідко це слово зустрічається у всіх інших документах колекції пошукової системи (IDF - inverse document frequency). Під колекцією пошукової системи розуміють всю сукупність документів, які доступні пошуковій системі. Вага документа за заданим пошуковим запитом розраховується як добуток цих двох параметрів.

Різні пошукові системи, крім параметрів TF і IDF, використовують додаткові коефіцієнти для обчислення ваги документа за заданим пошуковим запитом, але логіка розрахунку залишається незмінною: вага документа буде тим більше, чим частіше слово з пошукового запиту зустрічається в документі (до певних меж, після яких документ може бути визнано спамом) і чим рідше зустрічається це слово у всіх інших документах, проіндексованих пошуковою системою.

Методи штучного інтелекту

Швидкий розвиток сучасних технологій та постійне вдосконалення алгоритмів машинного навчання та “еволюція” розробки штучного інтелекту стає основою для публікацій та досліджень за цією тематикою. Розробники, науковці та дослідники приділяють увагу цим технологіям і визнають їхній значний вплив на майбутнє людства.

В роботах [19-21] запропонований інтелектуальний агент (замість агента-людини), який виконуватиме ті самі пошукові операції, взаємодіючи з роботодавцем і агентами-координаторами пошуку роботи. Запропоноване рішення передбачає створення агентів претендента, пошуку роботи та роботодавця, які використовуватимуть нечіткі правила переваг для прийняття правильного рішення щодо отримання списку вакансій на основі критеріїв пошуку користувача, а також подаватимуть рейтинг роботодавця на основі відгуків, поданих колишніми та нинішніми працівниками. Усі відповідні результати впорядковано на основі динамічного розрахунку очікуваної корисності від найвищого до найнижчого та відображено як список пошуку роботи.

В роботі [36] наводяться методи оцінки впливу пошуку роботи в Інтернеті на ймовірність повторного працевлаштування та тривалість періодів безробіття. У дослідженні використовуються національні панельні набори даних з Німеччини (SOEP 2003-2007) та Південної Кореї (KLIPS 1996-2006) для оцінки моделей пробіту та Хаусмана-Тейлора IV впливу пошуку роботи в Інтернеті на ймовірність повторне працевлаштування. Дослідження також вивчає аналіз тривалості з метою оцінки впливу пошуку роботи в Інтернеті на тривалість безробіття. В даній роботі встановлено, що у Німеччині та Південній Кореї шукачі роботи, які користувалися Інтернетом, мали на 7,1 і 12,7 відсоткових пунктів вищу ймовірність, відповідно, бути повторно працевлаштованими протягом наступних 12 місяців. Крім того, шукачі роботи, які користувалися Інтернетом, мали меншу тривалість безробіття як у Німеччині, так і в Південній Кореї. Практичні наслідки За останнє десятиліття рівень проникнення Інтернету та використання Інтернету для пошуку роботи

різко зросли в усьому світі. Інтернет значно змінив процес подання заявки на роботу та покращив канали спілкування між роботодавцями та шукачами. Результати дослідження свідчать про те, що Інтернет є корисним і має бути частиною пошуку роботи. Це дослідження, яке використовує панелі даних для аналізу зв'язку між використанням Інтернету та результатами пошуку роботи. Таким чином, результати стійкі до проблем неоднорідності, що не спостерігається. Дослідження також розглядає питання ендогенності пошуку роботи в Інтернеті за допомогою моделі Хаусмана-Тейлора IV. Рівень проникнення Інтернету та використання Інтернету для пошуку роботи різко зросли в усьому світі. Інтернет значно змінив процес подання заявки на роботу та покращив канали спілкування між роботодавцями та шукачами. Результати дослідження свідчать про те, що Інтернет є корисним і має бути частиною пошуку роботи.

У роботі [41] пропонуються методики та заходи для оцінки якості рейтингів пошуку роботи з точки зору моделювання користувачів. Користувачі веб-сайтів онлайн-пошуку роботи взаємодіють із ранжованими списками вакансій, підсумки, створені у відповідь на запити, сподіваючись визначити один або більше цікавих вакансій. Отже, якість рейтингу пошуку роботи стає основним фактором, який впливає на задоволеність користувачів. Досліджується поведінка шукачів роботи, коли вони взаємодіють зі сформованими рейтингами, використовуючи журнали взаємодії пошуку роботи з Seek.com, відомого австралійського веб-сайту пошуку роботи. Дані дослідження є основою для побудови моделі шукачів роботи які будуть включені в показник ефективності.

Останні пропозиції щодо моделей рейтингу пошуку роботи використовувалися з використанням двох типів показників для оцінки якості створеного рейтингу за моделями: (1) офлайн-метрики, такі як $NDCG@k$ (k встановлено як кількість резюме роботи, показаних на першій сторінці), $Prec@1$ або Mean взаємний ранг (MRR); і (2) онлайн-метрики, такі як кліки ставка та кількість заявок на роботу [43, 45].

В роботі [46] базуючись на визначенні пошуку роботи як цілеспрямованого, мотиваційного та саморегулюючого процесу, представлено структуру для впорядкування множини змінних, розглянутих у літературі щодо пошуку роботи та успішності працевлаштування. Проведений кількісний синтез літератури, щоб перевірити зв'язок між саморегуляцією під час пошуку роботи, поведінкою під час пошуку роботи та результатами успіху при працевлаштуванні. Також кількісно проаналізовано ключові передумови (наприклад, особистість, фактори ставлення та контекстуальні змінні) саморегуляції при пошуку роботи, поведінки під час пошуку роботи та успішності працевлаштування. Наведені дослідження, які вивчали зв'язки зі змінними успіху пошуку роботи чи працевлаштуванням серед шукачів (наприклад, нових учасників ринку праці, безробітних, зайнятих осіб), у результаті чого було отримано 378 незалежних вибірок ($N = 165\ 933$). Більшість вибірок (74,3%, $k = 281$) походять зі статей, опублікованих у 2001 році або пізніше. Результати метааналізу підтверджують роль інтенсивності пошуку роботи в прогнозуванні кількісних результатів успішного працевлаштування (тобто $r_c = 0,23$ для кількості співбесід, $r_c = 0,14$ для кількості пропозицій роботи та $r_c = 0,19$ для статусу зайнятості). Загальна інтенсивність пошуку роботи не змогла передбачити якість зайнятості. Зроблені висновки визначають саморегуляцію пошуку роботи та якість пошуку роботи як перспективні конструкції для майбутніх досліджень, оскільки вони передбачають як кількісні результати успішного працевлаштування, так і якість зайнятості. На основі результатів теоретичного та кількісного синтезу сформована програма майбутніх досліджень. (Запис бази даних PsycInfo (c) 2021 APA, усі права захищено). Досліджений зв'язок поведінки під час пошуку роботи з результатами успішного працевлаштування за наступними показниками.

1. Загальна інтенсивність пошуку роботи

1.1 Кількість співбесід

1.2 Кількість пропозицій роботи

- 1.3 Зайнятість
- 1.4 Якість працевлаштування
- 2. Активний пошук роботи**
- 2.1 Кількість співбесід
- 2.2 Кількість пропозицій роботи
- 2.3 Зайнятість
- 2.4 Якість працевлаштування
- 3. Підготовчий пошук роботи**
- 3.1 Кількість співбесід
- 3.2 Кількість пропозицій роботи
- 3.3 Зайнятість
- 3.4 Якість працевлаштування
- 4. Неформальний пошук роботи**
- 4.1 Кількість співбесід
- 4.2 Кількість пропозицій роботи
- 4.3 Зайнятість
- 4.4 Якість працевлаштування
- 5. Офіційний пошук роботи**
- 5.1 Кількість співбесід
- 5.2 Кількість пропозицій роботи
- 5.3 Зайнятість
- 5.4 Якість працевлаштування
- 6. Якість пошуку роботи**
- 6.1 Кількість співбесід
- 6.2 Кількість пропозицій роботи
- 6.3 Зайнятість
- 6.4 Якість працевлаштування

Взаємозв'язок саморегуляції під час пошуку роботи з поведінкою під час пошуку роботи та результатами успішного працевлаштування

1. Загальна саморегуляція пошуку роботи

- 1.1 Інтенсивність пошуку роботи
- 1.2 Якість пошуку роботи
- 1.3 Кількість співбесід
- 1.4 Кількість пропозицій роботи
- 1.5 Статус зайнятості
- 1.6 Якість працевлаштування
- 2. Ціль розвідки**
 - 2.1 Інтенсивність пошуку роботи
 - 2.2 Якість пошуку роботи
 - 2.3 Кількість співбесід
 - 2.4 Кількість пропозицій роботи
 - 2.5 Зайнятість
 - 2.6 Якість працевлаштування
- 3. Цільова чіткість**
 - 3.1 Інтенсивність пошуку роботи
 - 3.2 Якість пошуку роботи
 - 3.3 Кількість співбесід
 - 3.4 Кількість пропозицій роботи
 - 3.5 Зайнятість
 - 3.6 Якість працевлаштування
- 4. Наміри пошуку роботи**
 - 4.1 Інтенсивність пошуку роботи
 - 4.2 Якість пошуку роботи
 - 4.3 Кількість співбесід
 - 4.4 Кількість пропозицій роботи
 - 4.5 Зайнятість
 - 4.6 Якість працевлаштування
- 5. Акти саморегулювання**
 - 5.1 Інтенсивність пошуку роботи
 - 5.2 Якість пошуку роботи

- 5.3 Кількість співбесід
- 5.4 Кількість пропозицій роботи
- 5.5 Зайнятість
- 5.6 Якість працевлаштування

Методи статистичного та машинного навчання, Data Science

Аналіз останніх досліджень та публікацій дозволив виявити наступні проблеми автоматизації процесів пошуку роботи або кандидатів на певну посаду.

- Великі витрати часу в процесі пошуку роботи або кандидатів на певну посаду у зв'язку з недостатньою автоматизацією надання рекомендацій на різних етапах з метою зменшення області пошуку;
- Недостатній рівень задоволеності результатами пошуку, у зв'язку з чим значна частина потенційних користувачів звертається за рекомендаціями друзів, знайомих;
- Відсутність уніфікованої системи, яка допомогла б описати наявні професійні навички або вимоги до навичок кандидата, якого шукає роботодавець;
- Наявність різних підходів до оцінки ефективності систем пошуку, зокрема показників, які досліджуються при побудові відповідних моделей.

1.4 Постановка задачі

Провести аналіз існуючих моделей аналізу ефективності працевлаштування з урахуванням наборів характеристик процесу пошуку роботи на сайті.

Розділ 2 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ АНАЛІЗУ ЕФЕКТИВНОСТІ СИСТЕМ ПОШУКУ РОБОТИ МЕТОДАМИ DATA SCIENCE

2.1 Розробка методів формування наборів даних для аналізу ефективності систем пошуку роботи

Набори даних пошуку роботи

Аналіз характеристик наборів даних виконаємо на основі датасетів, що знаходяться у відкритому доступі:

- 1) <https://www.kaggle.com/datasets>
- 2) <https://github.com/>
- 3) <https://www.statista.com/>
- 4) PWC, Deloitte, KPMG консалтингові компанії які публікують свої дослідження у вільному доступі.
- 5) <https://stats.oarc.ucla.edu/ucla/books/#Logistic%20Regression%20and%20Related%20Methods>
- 6) <http://www.stata-press.com/data/r11/>
- 7) <https://r-data.pmagonia.com>
- 8) <https://github.com/vincentarelbundock/Rdatasets>

Показники для пошуку роботи:

1. регіон
2. тип зайнятості
3. галузь/вид діяльності
4. компанія
5. дата публікації оголошення роботодавцем
6. мова
7. заробітна плата середня
8. заробітна плата мінімальна
9. стаж роботи

10. рівень посади
11. тип роботи (он-лайн/ оф-лайн)
12. вік претендента мінімальний
13. вік претендента максимальний
14. режим роботи
15. термінове оголошення,
16. гаряча пропозиція,
17. кольоровий блок в оголошенні,
18. VSP-компанія,
19. автопоновлення оголошення,
20. платні публікації,
21. банерна реклама в оголошенні

Відгук Y - показники ефективності оголошення про вакансію:

8. Закрита/ не закрита вакансія за певний період (тиждень/місяць)
9. Час від появи оголошення до закриття вакансії
10. Кількість заяв на вакансію від претендентів за певний період (тиждень/місяць)

Множина можливих предикторів X :

Кількісні характеристики процесу пошуку роботи на сайті:

1. Дата публікації оголошення роботодавцем
2. Заробітна плата середня
3. Заробітна плата мінімальна
4. Вік претендента мінімальний
5. Вік претендента максимальний

Характеристики компанії:

1. Рейтинг компанії
2. Розмір компанії
3. Рік заснування
4. Локація
5. Назва компанії

Якісні характеристики процесу пошуку роботи на сайті:

Характеристики виду оголошення

1. термінове оголошення,
2. гаряча пропозиція,
3. кольоровий блок в оголошенні,
4. VSP-компанія,
5. автопоновлення оголошення,
6. платні публікації,
7. банерна реклама в оголошенні

Відгук Z - показники ефективності резюме:

1. Влаштувався/не влаштувався пошукач на роботу
2. Перша заробітна плата.
3. Різниця між першою заробітною платою та заявленою у резюме.

Характеристики претендента на роботу

1. регіон
2. галузь/вид діяльності
3. компанія
4. мова
5. заробітна плата очікувана
6. заробітна плата мінімальна
7. стаж роботи
8. рівень посади
9. вік претендента

Характеристики умов праці

1. тип зайнятості (повний/частковий)
2. тип роботи (он-лайн/ оф-лайн)
3. режим роботи

Завдяки засобам штучного інтелекту можливо ефективно вирішити проблему скорочення часу пошуку вакансії або кандидата.

Спираючись на те, за якими саме критеріями користувач опрацьовує пошук, системи автоматично пропонує вакансії, які можуть бути цікавими для користувача. Одночасно, аналізуючи інформацію про успішно укомплектовані вакансії, система підлаштовується самостійно і змінює так зване коло кандидатів, резюме яких, найімовірніше, відповідають умовам подібних вакансій.



4 РОЗРОБКА ПРОГРАМНОГО ЗАСОБУ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

Розробку програмного забезпечення продемонструємо на прикладі аналізу даних про працевлаштування <https://r-data.pmagunia.com>, <https://github.com/vincentarelbundock/Rdatasets> за допомогою Google Collab засобами мови R.

The screenshot displays a Google Colab environment with the following R code and output:

```
ds_salaries = read.csv("ds_salaries.csv", header = T, na.strings = "?")
ds_salaries [1:5,]
Salaries=ds_salaries
```

Top Data Frame (5 x 12):

X	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
<int>	<int>	<chr>	<chr>	<chr>	<int>	<chr>	<int>	<chr>	<int>	<chr>	<chr>
1	0	2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0	DE
2	1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0	JP
3	2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50	GB
4	3	2020	MI	FT	Product Data Analyst	20000	USD	20000	HN	0	HN
5	4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	US	50	US

Bottom Data Frame (5 x 12):

work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
<int>	<chr>	<chr>	<chr>	<int>	<chr>	<int>	<chr>	<int>	<chr>	<chr>
2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0	DE	L
2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0	JP	S
2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50	GB	M
2020	MI	FT	Product Data Analyst	20000	USD	20000	HN	0	HN	S
2020	SE	FT	Machine Learning Engineer	150000	USD	150000	US	50	US	L

Рис. 3.1 Фрагменти набору даних характеристик вакансій

```

Отримання загальних характеристик набору даних. Розмір, назви стовпців(показників), описові статистики показників
#розмір
dim(Salaries)
#назви показників
names(Salaries)
#описові статистики показників
summary(Salaries)

607 - 12
'X' 'work_year' 'experience_level' 'employment_type' 'job_title' 'salary' 'salary_currency' 'salary_in_usd' 'employee_residence' 'remote_ratio' 'company_location' 'company_size'
  X          work_year  experience_level  employment_type
Min.   : 0.0   Min.   :2020   Length:607     Length:607
1st Qu.:151.5 1st Qu.:2021   Class :character Class :character
Median :303.0 Median :2022   Mode  :character Mode  :character
Mean   :303.0 Mean   :2021
3rd Qu.:454.5 3rd Qu.:2022
Max.   :606.0 Max.   :2022

  job_title      salary      salary_currency  salary_in_usd
Length:607      Min.   : 4000   Length:607      Min.   : 2859
Class :character 1st Qu.: 70000   Class :character 1st Qu.: 62726
Mode  :character Median : 115000   Mode  :character Median :101570
                          Mean  : 324000                          Mean  :112298
                          3rd Qu.: 165000                      3rd Qu.:150000
                          Max.   :30400000                     Max.   :600000

  X          work_year  experience_level  employment_type
Min.   : 0.0   Min.   :2020   Length:607     Length:607
1st Qu.:151.5 1st Qu.:2021   Class :character Class :character
Median :303.0 Median :2022   Mode  :character Mode  :character
Mean   :303.0 Mean   :2021
3rd Qu.:454.5 3rd Qu.:2022
Max.   :606.0 Max.   :2022

  job_title      salary      salary_currency  salary_in_usd
Length:607      Min.   : 4000   Length:607      Min.   : 2859
Class :character 1st Qu.: 70000   Class :character 1st Qu.: 62726
Mode  :character Median : 115000   Mode  :character Median :101570
                          Mean  : 324000                          Mean  :112298
                          3rd Qu.: 165000                      3rd Qu.:150000
                          Max.   :30400000                     Max.   :600000

  employee_residence  remote_ratio  company_location  company_size
Length:607           Min.   : 0.00   Length:607       Length:607
Class :character     1st Qu.: 50.00   Class :character Class :character
Mode  :character     Median :100.00   Mode  :character Mode  :character
                          Mean   : 70.92
                          3rd Qu.:100.00
                          Max.   :100.00

```

Рис. 3.2 Характеристики набору даних та окремих показників

```

▶ par(mfrow = c(2, 2))
  hist(salary, col="red")
  hist(salary_in_usd, col="green")
  hist(remote_ratio, col="blue")

```

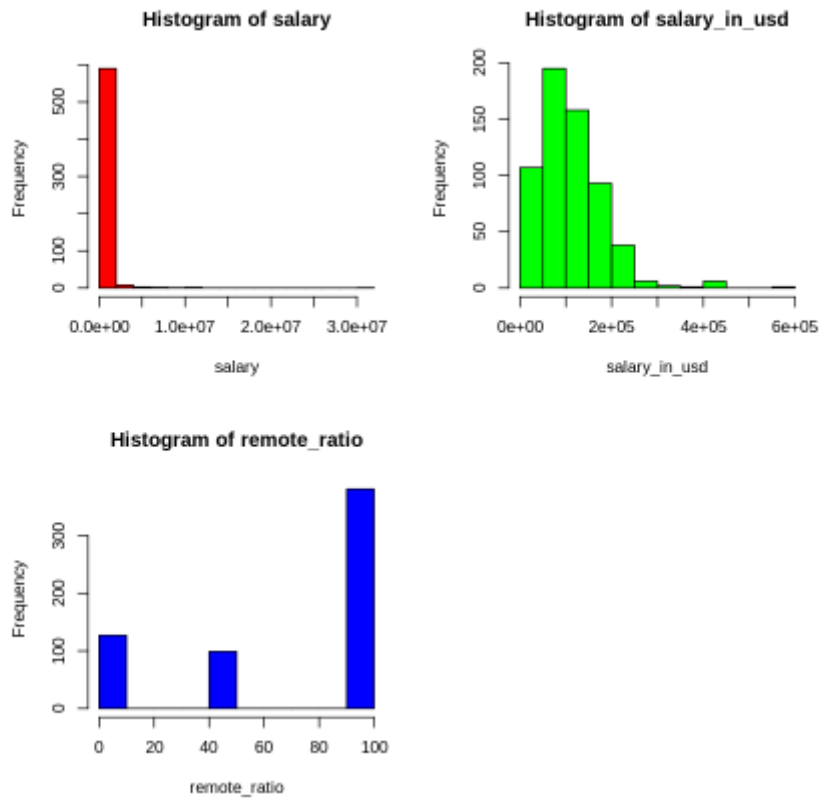

0
лек.

Рис. 3.3 Аналіз статистичних розподілів показників

Побудова діаграм розсіювання показників

```

▶ work_year=as.factor(work_year)
  x=work_year
  y=salary
  plot(x, y,varwidth = T, col ="green", xlab = "work_year", ylab="salary")

```

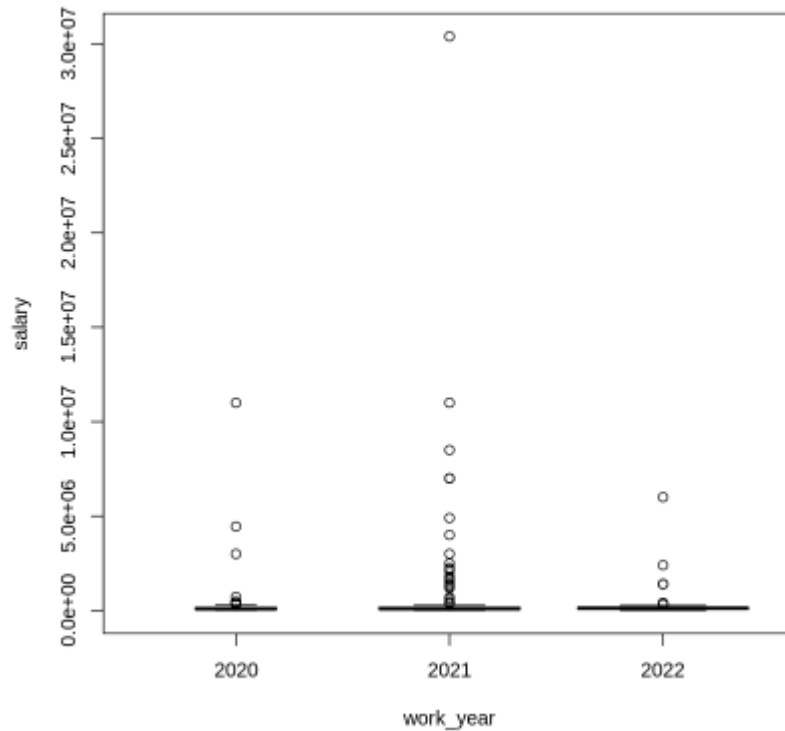


Рис. 3.4 Діаграма розсіювання показника salary в залежності від work_year

```
summary(work_year)
```

```
2020:    72 2021:   217 2022:   318
```

```
experience_level=as.factor(experience_level)
x=experience_level
y=salary
plot(x, y,varwidth = T, xlab = "experience_level", ylab="salary")
```

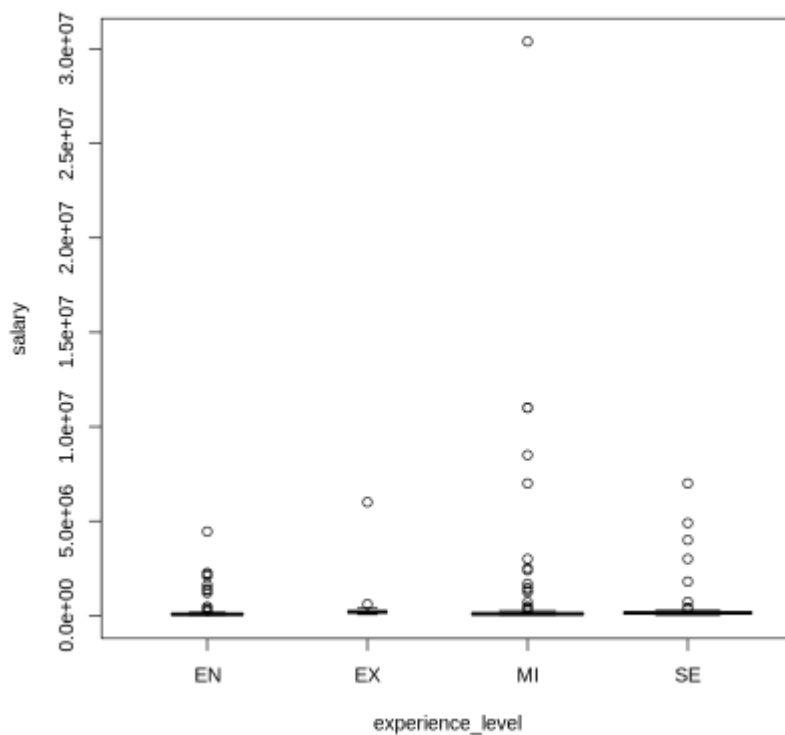


Рис. 3.5 Діаграма розсіювання показника salary в залежності від experience_level

```
summary(experience_level)
```

```
EN:      88 EX:      26 MI:     213 SE:     280
```

```
employment_type =as.factor(employment_type)
```

```
x=employment_type
```

```
y=salary
```

```
plot(x, y,varwidth = T, xlab = "employment_type", ylab="salary")
```

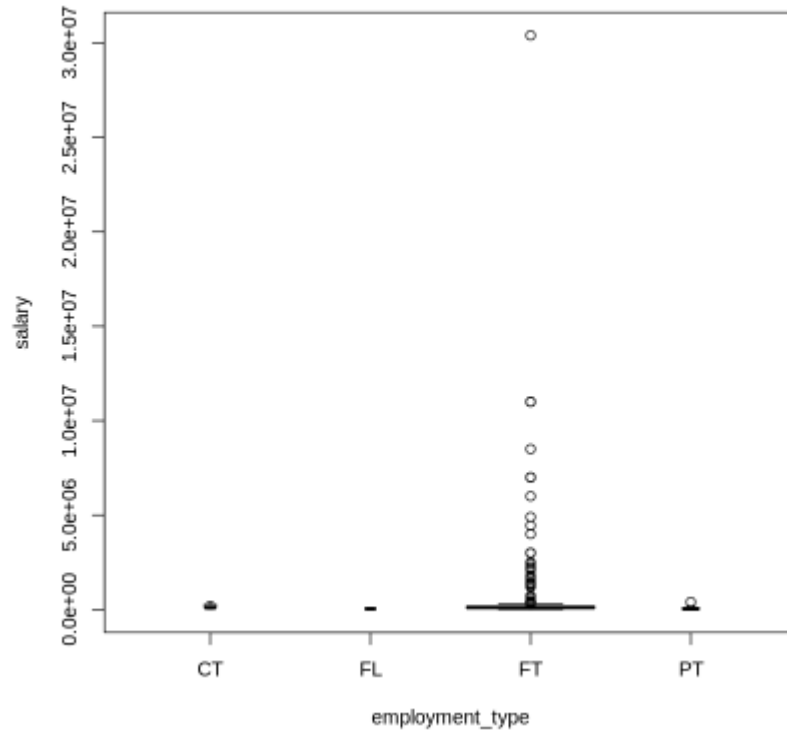


Рис. 3.6 Діаграма розсіювання показника salary в залежності від employment_type

```
remote_ratio = as.factor(remote_ratio)
x=remote_ratio
y=salary
plot(x, y, varwidth = T, xlab = "remote_ratio", ylab="salary")
```

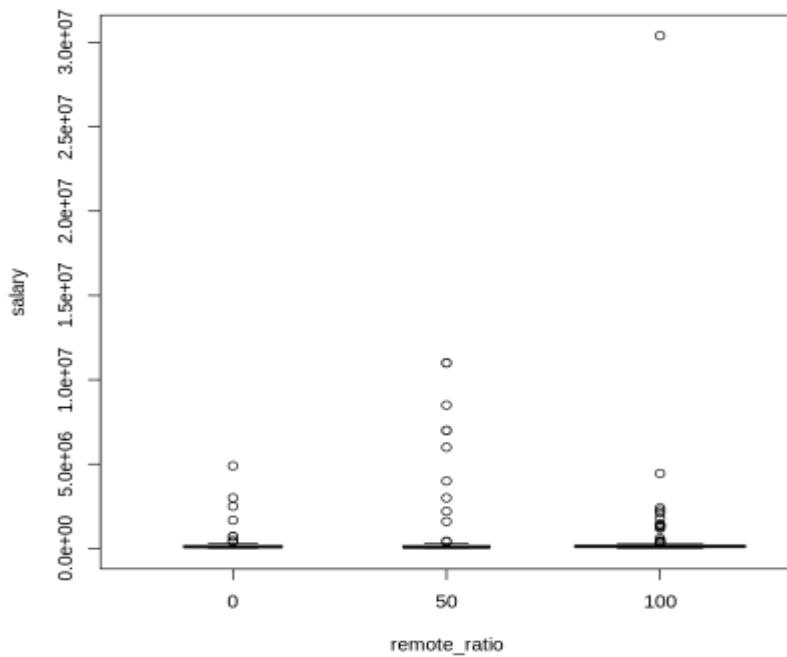


Рис. 3.7 Діаграма розсіювання показника salary в залежності від remote_ratio

```
summary(remote_ratio)
```

```
0:      127 50:      99 100:      381
```

Видалення максимального значення.

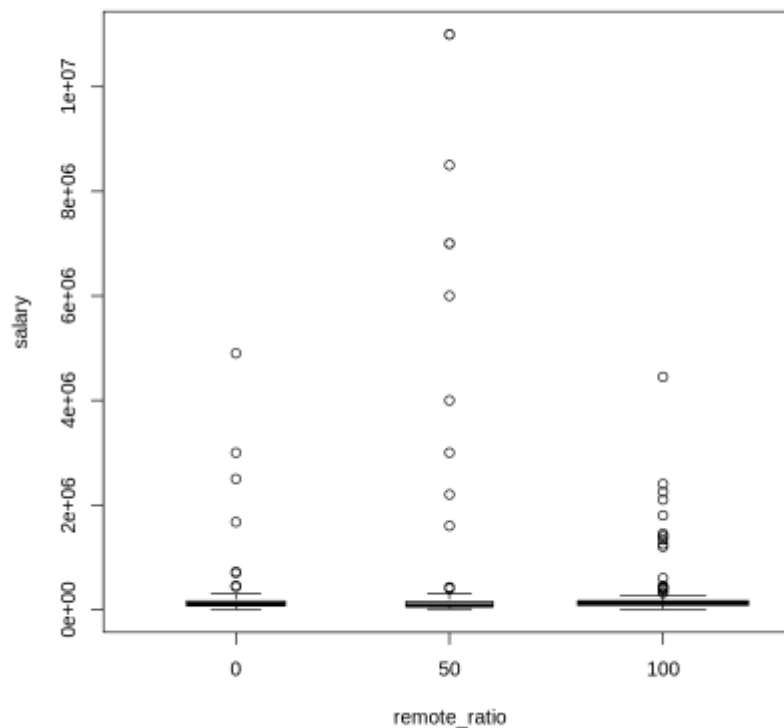
```
max(Salaries$salary)
summary(Salaries$salary)
Salaries$salary [Salaries$salary == 30400000] <- NaN
summary(Salaries$salary)
```

```
30400000
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4000	70000	115000	324000	165000	30400000

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
4000	70000	114000	274370	164999	11000000	1

```
x=remote_ratio
y=salary
plot(x, y,varwidth = T, xlab = "remote_ratio", ylab="salary")
```



```
▶ attach(Salaries)
```

```
↳ The following objects are masked _by_ .GlobalEnv:
```

```
employment_type, experience_level, remote_ratio, work_year
```

```
The following objects are masked from Salaries (pos = 3):
```

```
company_location, company_size, employee_residence,  
employment_type, experience_level, job_title, remote_ratio, salary,  
salary_currency, salary_in_usd, work_year, X
```

```
▶ Salaries$salary [Salaries$salary > 1000000] <- NaN  
summary(Salaries$salary)
```

```
↳
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
4000	67000	110000	123685	157000	720000	26

```
▶ x=remote_ratio  
y=salary  
plot(x, y,varwidth = T, xlab = "remote_ratio", ylab="salary")
```

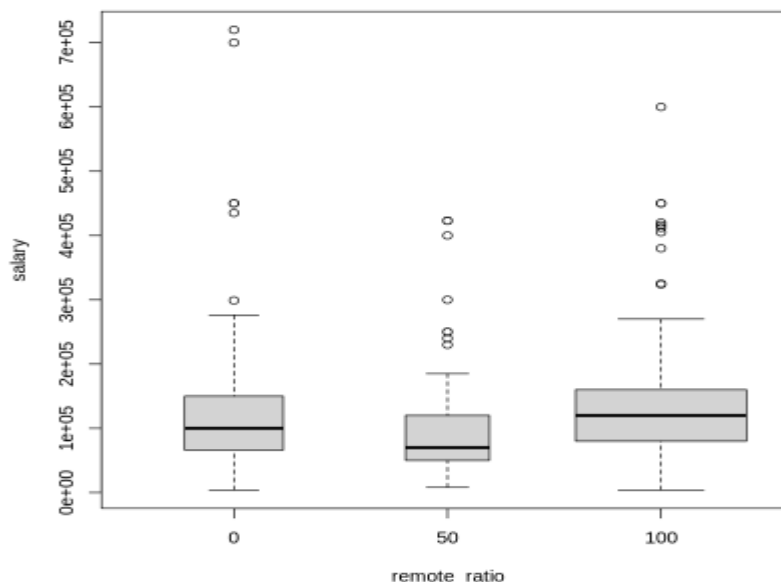


Рис. 3.9 Діаграма розсіювання показника salary в залежності від remote_ratio (після видалення аномально великих значень)

```
summary(remote_ratio)
```

```
0:      127 50:      99 100:      381
```

```
▶ work_year=as.factor(work_year)
  x=work_year
  y=salary
  plot(x, y,varwidth = T, col ="green", xlab = "work_year", ylab="salary")
```

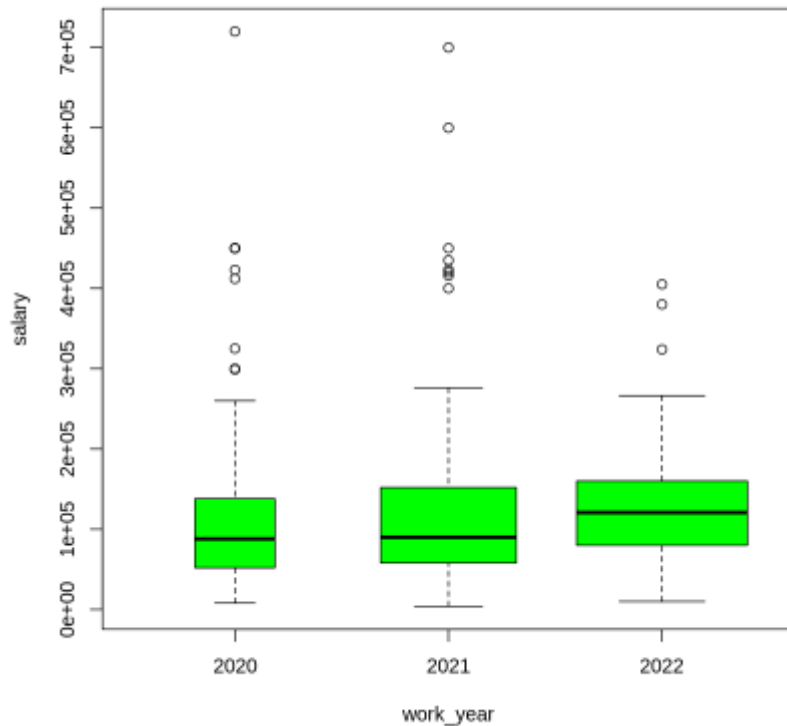


Рис. 3.10 Діаграма розсіювання показника salary в залежності від work_year (після видалення аномально великих значень)

```
summary(work_year)
```

```
2020:      72 2021:      217 2022:      318
```

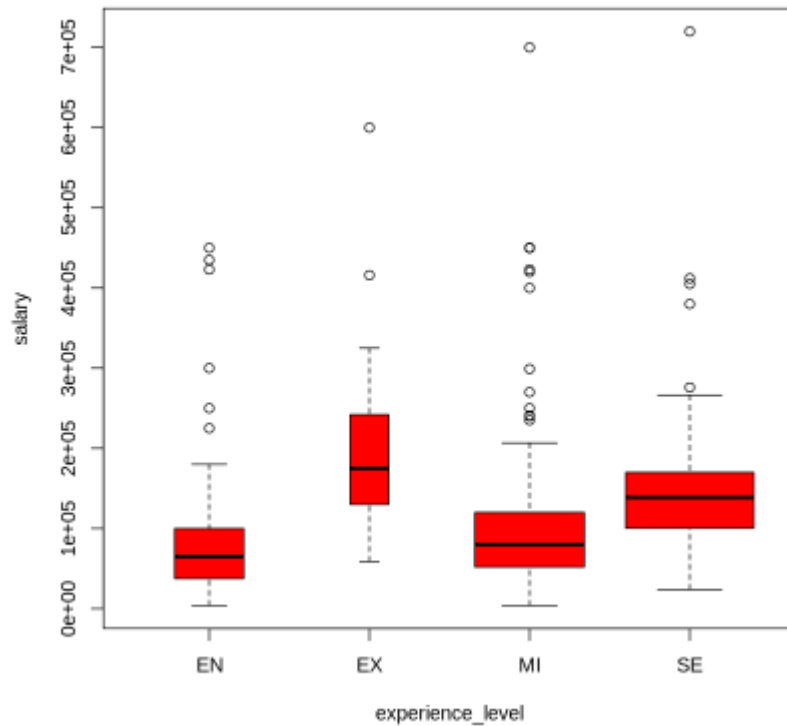


Рис. 3.11 Діаграма розсіювання показника salary в залежності від experience_level (після видалення аномально великих значень)

```
summary(experience_level)
```

```
EN:      88 EX:      26 MI:     213 SE:     280
```

```
▶ employment_type = as.factor(employment_type)
x=employment_type
y=salary
plot(x, y, varwidth = T, col = "blue", xlab = "employment_type", ylab = "salary")
```

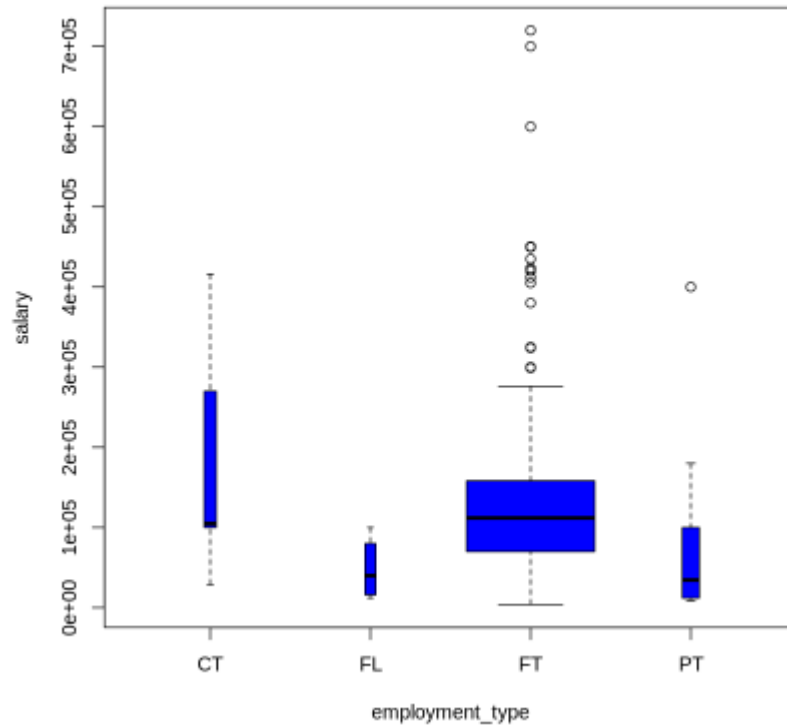



Рис. 3.12 Діаграма розсіювання показника salary в залежності від employment_type (після видалення аномально великих значень)

```
summary(employment_type)
```

```
CT:      5 FL:      4 FT:    588 PT:     10
```

```
par(mfrow = c(2, 2))
x=remote_ratio
y=salary
plot(x, y,varwidth = T, xlab = "remote_ratio", ylab="salary")
x=work_year
y=salary
plot(x, y,varwidth = T, xlab = "work_year", ylab="salary")
x=experience_level
y=salary
plot(x, y,varwidth = T, xlab = "experience_level", ylab="salary")
x=employment_type
y=salary
plot(x, y,varwidth = T, xlab = "employment_type", ylab="salary")
```

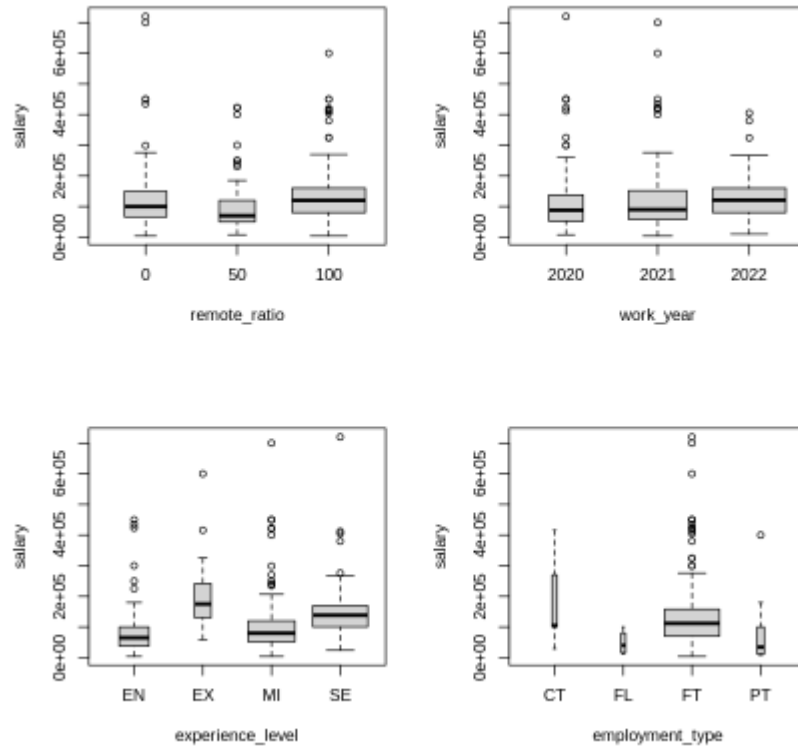


Рис. 3.13 Діаграми розсіювання показника salary в залежності від набору факторів (після видалення аномально великих значень)

```

▶ job_title =as.factor(job_title)
x=job_title
y=salary
plot(x, y,varwidth = T, col ="blue", xlab = "job_title", ylab="salary")

```

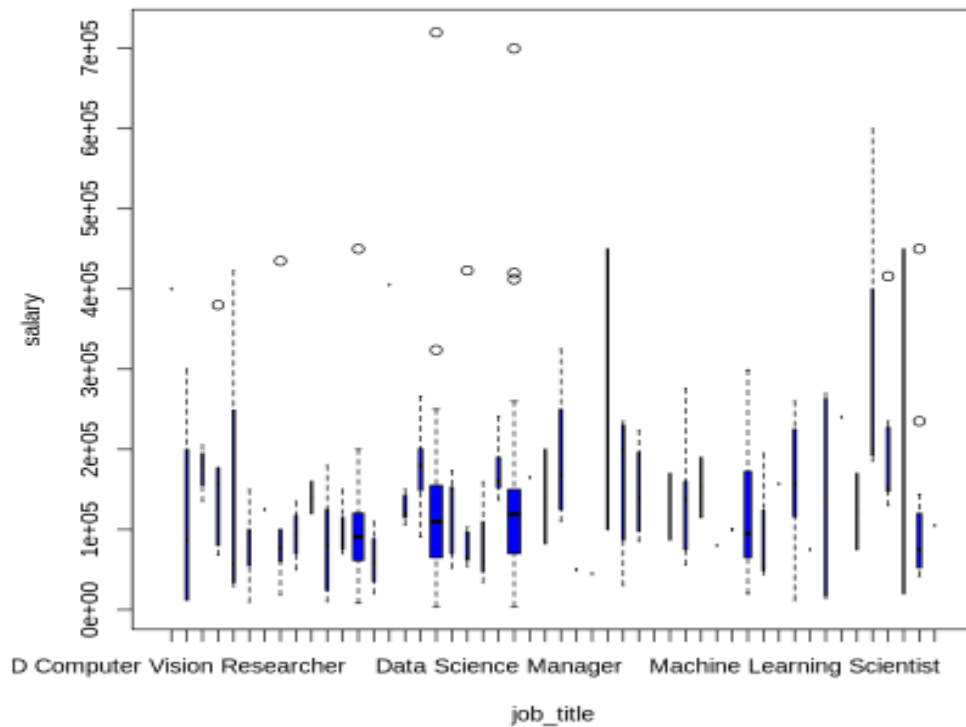


Рис. 3.14 Діаграми розсіювання показника salary в залежності від фактора job_title

summary(job_title)

```

3D Computer Vision Researcher: 1 AI Scientist: 7 Analytics Engineer: 4 Applied Data Scientist: 5 Applied Machine Learning Scientist: 4 BI Data Analyst: 6
Big Data Architect: 1 Big Data Engineer: 8 Business Data Analyst: 5 Cloud Data Engineer: 2 Computer Vision Engineer: 6 Computer Vision Software Engineer: 3
Data Analyst: 97 Data Analytics Engineer: 4 Data Analytics Lead: 1 Data Analytics Manager: 7 Data Architect: 11 Data Engineer: 132 Data Engineering Manager:
5 Data Science Consultant: 7 Data Science Engineer: 3 Data Science Manager: 12 Data Scientist: 143 Data Specialist: 1 Director of Data Engineering: 2
Director of Data Science: 7 ETL Developer: 2 Finance Data Analyst: 1 Financial Data Analyst: 2 Head of Data: 5 Head of Data Science: 4
Head of Machine Learning: 1 Lead Data Analyst: 3 Lead Data Engineer: 6 Lead Data Scientist: 3 Lead Machine Learning Engineer: 1 Machine Learning Developer:
3 Machine Learning Engineer: 41 Machine Learning Infrastructure Engineer: 3 Machine Learning Manager: 1 Machine Learning Scientist: 8 Marketing Data Analyst:
1 ML Engineer: 6 NLP Engineer: 1 Principal Data Analyst: 2 Principal Data Engineer: 3 Principal Data Scientist: 7 Product Data Analyst: 2
Research Scientist: 16 Staff Data Scientist: 1

```

```

salary_currency =as.factor(salary_currency)
x=salary_currency
y=salary
plot(x, y,varwidth = T, col ="blue", xlab = "salary_currency", ylab="salary")

```

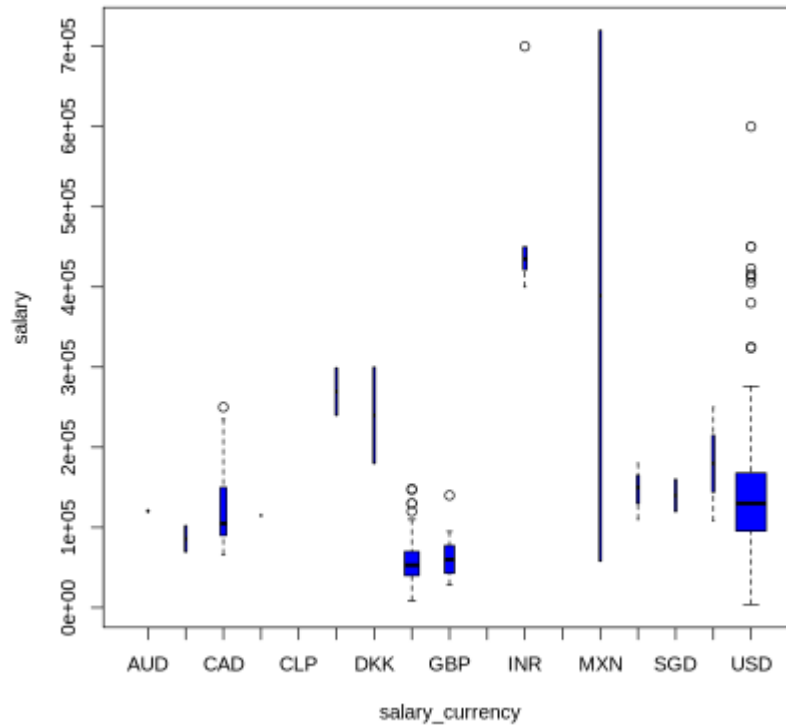


Рис. 3.15 Діаграми розсіювання показника salary в залежності від фактора salary_currency

```
summary(salary_currency)
```

```
AUD: 2 BRL: 2 CAD: 18 CHF: 1 CLP: 1 CNY: 2 DKK: 2 EUR: 95 GBP: 44 HUF: 2 INR: 27 JPY: 3 MXN: 2 PLN: 3 SGD: 2 TRY: 3 USD: 398
```

```
▶ employee_residence =as.factor(employee_residence)
x=employee_residence
y=salary
plot(x, y,varwidth = T, col ="blue", xlab = "employee_residence", ylab="salary")
```

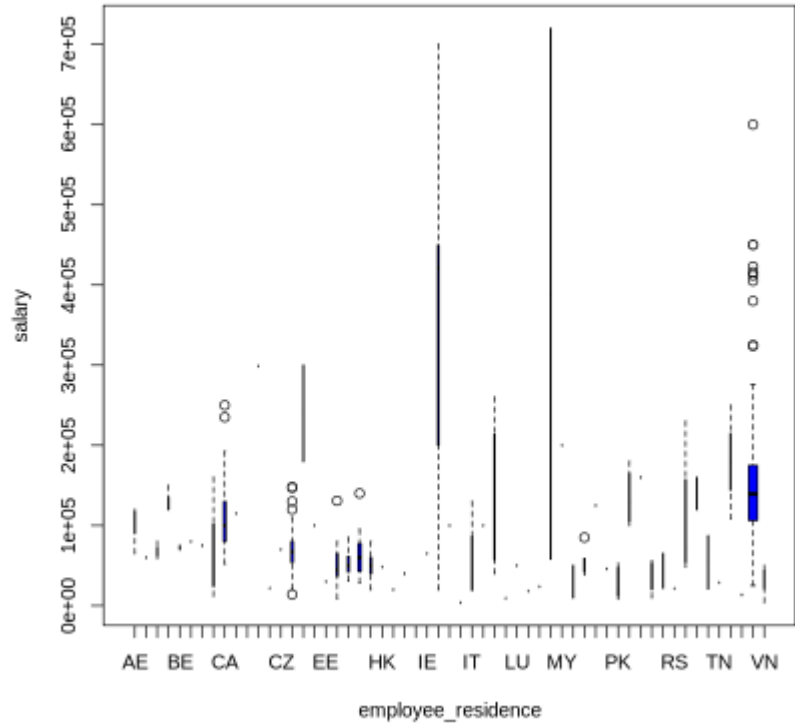


Рис. 3.16 Діаграми розсіювання показника salary в залежності від фактора employee_residence

```
summary(employee_residence)
```

AE:	3 AR:	1 AT:	3 AU:	3 BE:	2 BG:	1 BO:	1 BR:	6 CA:	29 CH:	1 CL:	1 CN:	1 CO:	1 CZ:	1 DE:	25 DK:	
2 DZ:	1 EE:	1 ES:	15 FR:	18 GB:	44 GR:	13 HK:	1 HN:	1 HR:	1 HU:	2 IE:	1 IN:	30 IQ:	1 IR:	1 IT:	4	
JE:	1 JP:	7 KE:	1 LU:	1 MD:	1 MT:	1 MX:	2 MY:	1 NG:	2 NL:	5 NZ:	1 PH:	1 PK:	6 PL:	4 PR:	1 PT:	6
RO:	2 RS:	1 RU:	4 SG:	2 SI:	2 TN:	1 TR:	3 UA:	1 US:	332 VN:	3						

Побудова моделей для оцінки і прогнозу ефективності від наборів факторів, що характеризують претендента на роботу.

```
lm(formula=salary_in_usd~employment_type+experience_level+remote_ratio+company_size, data = Salaries)  
summary(lm(formula=salary_in_usd~employment_type+experience_level+remote_ratio+company_size, data = Salaries))
```

Call:

```
lm(formula = salary_in_usd ~ employment_type + experience_level +
    remote_ratio + company_size, data = Salaries)
```

Coefficients:

(Intercept)	employment_typeFL	employment_typeFT	employment_typePT
137534.1	-119535.2	-70405.4	-104171.0
experience_levelEX	experience_levelMI	experience_levelSE	remote_ratio
128044.2	22730.6	71622.1	130.5
company_sizeM	company_sizeS		
-11815.0	-28796.5		

Residuals:

Min	1Q	Median	3Q	Max
-127456	-39324	-5613	28260	391780

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	137534.09	28366.65	4.848	1.59e-06	***
employment_typeFL	-119535.22	40925.65	-2.921	0.003623	**
employment_typeFT	-70405.36	27413.66	-2.568	0.010463	*
employment_typePT	-104170.96	33401.36	-3.119	0.001904	**
experience_levelEX	128044.23	13712.17	9.338	< 2e-16	***
experience_levelMI	22730.61	7940.93	2.862	0.004351	**
experience_levelSE	71622.08	7871.82	9.099	< 2e-16	***
remote_ratio	130.47	61.19	2.132	0.033393	*
company_sizeM	-11814.97	5563.46	-2.124	0.034108	*
company_sizeS	-28796.47	8074.74	-3.566	0.000391	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60580 on 597 degrees of freedom

Multiple R-squared: 0.2819, Adjusted R-squared: 0.2711

F-statistic: 26.04 on 9 and 597 DF, p-value: < 2.2e-16

```
lm.fit = lm(salary_in_usd ~ employment_type + experience_level +
    company_size+job_title, data = Salaries)
summary(lm.fit)
```

Call:

```
lm(formula = salary_in_usd ~ employment_type + experience_level +
    company_size + job_title, data = Salaries)
```

Residuals:

Min	1Q	Median	3Q	Max
-164314	-31712	-2982	24742	348153

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	84633.8	66957.1	1.264
employment_typeFL	-131281.5	42624.5	-3.080
employment_typeFT	-74457.7	31162.0	-2.389
employment_typePT	-99771.6	36215.2	-2.755
experience_levelEX	125078.5	16003.2	7.816
experience_levelMI	21372.3	7592.9	2.815
experience_levelSE	68859.1	7695.9	8.948
company_sizeM	-825.5	5652.8	-0.146
company_sizeS	-23960.0	7903.1	-3.032
job_titleAI Scientist	57753.0	61424.1	0.940
job_titleAnalytics Engineer	68680.6	65648.0	1.046
job_titleApplied Data Scientist	129386.3	64322.7	2.012
job_titleApplied Machine Learning Scientist	97661.7	65623.0	1.488
job_titleBI Data Analyst	41308.3	63487.1	0.651
job_titleDirector of Data Science	71391.5	64125.0	1.113
job_titleETL Developer	24234.1	70761.6	0.342
job_titleFinance Data Analyst	-17139.2	81282.5	-0.211
job_titleFinancial Data Analyst	254137.7	71256.4	3.567
job_titleHead of Data	73094.1	64465.4	1.134
job_titleHead of Data Science	43999.8	66162.6	0.665
job_titleHead of Machine Learning	-56215.6	82529.1	-0.681
job_titleLead Data Analyst	44825.6	67313.4	0.666
job_titleLead Data Engineer	75344.8	63618.3	1.184
job_titleLead Data Scientist	59970.4	67386.6	0.890
job_titleLead Machine Learning Engineer	9722.3	81021.1	0.120
job_titleMachine Learning Developer	54144.4	67315.8	0.804
job_titleMachine Learning Engineer	62193.7	59851.8	1.039
job_titleMachine Learning Infrastructure Engineer	54318.0	67085.9	0.810
job_titleMachine Learning Manager	38068.8	81282.5	0.468
job_titleMachine Learning Scientist	125027.4	62414.2	2.003
job_titleMarketing Data Analyst	9618.8	81282.5	0.118
job_titleML Engineer	92947.9	62766.6	1.481
job_titleNLP Engineer	5687.5	81084.1	0.070
job_titlePrincipal Data Analyst	79600.9	70948.9	1.122
job_titlePrincipal Data Engineer	230833.5	67508.9	3.419
job_titlePrincipal Data Scientist	128099.5	62865.9	2.038
job_titleProduct Data Analyst	-6532.5	71013.9	-0.092
job_titleResearch Scientist	71124.3	60857.6	1.169
job_titleStaff Data Scientist	-47667.4	86448.8	-0.551

```

(Intercept)                                0.206766
employment_typeFL                          0.002174 **
employment_typeFT                          0.017214 *
employment_typePT                          0.006065 **
experience_levelEX                         2.8e-14 ***
experience_levelMI                         0.005056 **
experience_levelSE                         < 2e-16 ***
company_sizeM                              0.883945
company_sizeS                              0.002546 **
job_titleAI Scientist                      0.347511
job_titleAnalytics Engineer               0.295931
job_titleApplied Data Scientist           0.044759 *
job_titleApplied Machine Learning Scientist 0.137266
job_titleBI Data Analyst                  0.515540
job_titleBig Data Architect                0.790893
job_titleBig Data Engineer                0.715697
job_titleBusiness Data Analyst            0.545936
job_titleCloud Data Engineer              0.252968
job_titleComputer Vision Engineer         0.725563
job_titleComputer Vision Software Engineer 0.124246
job_titleData Analyst                     0.518404
job_titleLead Data Engineer               0.236797
job_titleLead Data Scientist              0.373885
job_titleLead Machine Learning Engineer  0.904529
job_titleMachine Learning Developer       0.421552
job_titleMachine Learning Engineer        0.299202
job_titleMachine Learning Infrastructure Engineer 0.418476
job_titleMachine Learning Manager         0.639719
job_titleMachine Learning Scientist       0.045648 *
job_titleMarketing Data Analyst           0.905843
job_titleML Engineer                      0.139221
job_titleNLP Engineer                     0.944105
job_titlePrincipal Data Analyst           0.262375
job_titlePrincipal Data Engineer          0.000674 ***
job_titlePrincipal Data Scientist         0.042063 *
job_titleProduct Data Analyst             0.926741
job_titleResearch Scientist               0.243031
job_titleStaff Data Scientist             0.581588

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 55420 on 549 degrees of freedom
Multiple R-squared:  0.4473,    Adjusted R-squared:  0.3899
F-statistic: 7.796 on 57 and 549 DF,  p-value: < 2.2e-16

```

```

▶ summary(lm(formula=salary_in_usd~., data = Salaries))

```



```

Call:
lm(formula = salary_in_usd ~ ., data = Salaries)

Residuals:
    Min       1Q   Median       3Q      Max
-157428  -20638         0   15269  255828

Coefficients: (34 not defined because of singularities)
              Estimate Std. Error t value
(Intercept)  3.629e+07  1.303e+07  2.784
X             4.264e+01  2.413e+01  1.767
work_year    -1.794e+04  6.451e+03  -2.781
experience_levelEX  1.153e+05  1.446e+04  7.973
experience_levelMI  2.627e+04  7.086e+03  3.707
experience_levelSE  4.548e+04  7.488e+03  6.074
employment_typeFL -1.094e+05  4.814e+04  -2.272
employment_typeFT -7.805e+04  2.905e+04  -2.687
employment_typePT -8.975e+04  3.408e+04  -2.634
job_titleAI Scientist  5.382e+03  5.479e+04  0.098
job_titleAnalytics Engineer -2.402e+04  5.508e+04  -0.436
job_titleApplied Data Scientist  6.758e+04  5.537e+04  1.221
.....
company_locationMT NA
company_locationMX NA
company_locationMY NA
company_locationNG NA
company_locationNL 0.002262 **
company_locationNZ NA
company_locationPK 0.014399 *
company_locationPL 0.028256 *
company_locationPT 0.011437 *
company_locationRO 0.014018 *
company_locationRU 0.027446 *
company_locationSG NA
company_locationSI NA
company_locationTR NA
company_locationUA NA
company_locationUS 0.005779 **
company_locationVN NA
company_sizeM 0.863205
company_sizeS 0.031586 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 43430 on 458 degrees of freedom
Multiple R-squared:  0.7169,    Adjusted R-squared:  0.6255
F-statistic: 7.838 on 148 and 458 DF,  p-value: < 2.2e-16

```

Рис. 3.16 Звіт за результатами оцінювання моделі ефективності
працевлаштування

Висновки до розділу 3

В даному розділі побудовані моделі оцінки ефективності у випадку, коли пошукач роботи вибирає в якості критерію розмір заробітної плати. Дані методи можуть бути поширені на випадок, коли в якості критерію обирається час пошуку роботи і також за наборами факторів що характеризують інтенсивність пошуку роботи на сайті, та інтереси пошукача роботи. Оціночні значення розраховані по моделях надалі використовуються в системі для формування рекомендацій (ранжування пропозицій), які можуть бути цікавими для користувача.

Також на базі подібних моделей, система аналізуючи інформацію про успішно укомплектовані вакансії, підлаштовується самостійно і змінює так зване коло кандидатів, резюме яких, найімовірніше, відповідають умовам подібних вакансій.

ВИСНОВКИ

Предметом дослідження даної роботи є методи Data Science для аналізу статистичних даних, які можуть використовуватися для оцінювання ефективності систем пошуку роботи.

Визначені показники, які характеризують працевлаштування для кандидата на роботу і роботодавця в автоматизованій системі пошуку роботи за різними наборами факторів, а саме.

Показники для пошуку роботи: регіон; тип зайнятості; галузь/вид діяльності; компанія; дата публікації оголошення роботодавцем; мова; заробітна плата середня; заробітна плата мінімальна; стаж роботи; рівень посади; тип роботи (он-лайн/ оф-лайн); вік претендента мінімальний; вік претендента максимальний; режим роботи; термінове оголошення; гаряча пропозиція; кольоровий блок в оголошенні; VSP-компанія; автопоновлення оголошення; платні публікації; банерна реклама в оголошенні.

Відгук Y - показники ефективності оголошення про вакансію: закрыта/ не закрыта вакансія за певний період (тиждень/місяць); час від появи оголошення до закриття вакансії; кількість заяв на вакансію від претендентів за певний період (тиждень/місяць)

Кількісні характеристики процесу пошуку роботи на сайті: дата публікації оголошення роботодавцем; заробітна плата середня; заробітна плата мінімальна; вік претендента мінімальний; вік претендента максимальний.

Характеристики компанії: рейтинг компанії; розмір компанії; рік заснування; локація; назва компанії.

Характеристики виду оголошення: термінове оголошення; гаряча пропозиція; кольоровий блок в оголошенні; VSP-компанія; автопоновлення оголошення; платні публікації; банерна реклама в оголошенні

Відгук Z - показники ефективності резюме: влаштувався/не влаштувався пошукач на роботу; перша заробітна плата; різниця між першою заробітною платою та заявленою у резюме.

Характеристики претендента на роботу: регіон; галузь/вид діяльності; мова; заробітна плата мінімальна; розмір надбавок; стаж роботи; рівень посади на попередньому місці роботи; вік претендента.

Характеристики умов праці: тип зайнятості (повний/частковий); тип роботи (он-лайн/ оф-лайн); режим роботи.

Встановлений перелік факторів що впливає на ефективність працевлаштування. Розроблено моделі і здійснена програмна реалізація прогнозу ефективності працевлаштування за можливими наборами факторів, що впливають на ефективність працевлаштування.



Література

1. Мінківська, М. В., Єпик, М. О. Веб-орієнтована система пошуку вакансій роботи. Матеріали всеукраїнської науково-практичної конференції «Комп'ютерні технології обробки даних». Вінниця, Донецький національний університет імені Василя Стуса. Кафедра інформаційних технологій. 04.12.2020 р. URL: <https://jktod.donnu.edu.ua/issue/view/369>, p. 100-102.
2. Barber, L. (2006), "E-recruitment developments", HR Network Paper MP63, Institute for Employment Studies, Brighton
3. Beard, R., Ford, G. and Saba, R. (2010), "Internet use and job search", Phoenix Center Policy Paper No. 39, The Phoenix Center for Advanced Legal & Economic Public Policy Studies, Washington, DC.
4. Work.ua [Електронний ресурс] – Режим доступу: <https://www.work.ua/> (дата звернення 20.11.2022р) – Назва з екрана.
5. Rabota.ua [Електронний ресурс] – Режим доступу: <http://www.rabota.ua/> (дата звернення 20.11.2022р) – Назва з екрана.
6. [Електронний ресурс] – Режим доступу: <https://jobs.ua/> (дата звернення 20.11.2022р) – Назва з екрана.
7. Dou.ua [Електронний ресурс] – Режим доступу: <https://dou.ua/> (дата звернення 20.11.2022р) – Назва з екрана.
8. Endeed [Електронний ресурс] – Режим доступу: <https://ua.endeed.com/> (дата звернення 20.11.2022р) – Назва з екрана.
9. LinkedIn [Електронний ресурс] – Режим доступу: <https://https://www.linkedin.com/> (дата звернення 20.11.2022р) – Назва з екрана.
10. Джинні [Електронний ресурс] – Режим доступу: <https://djinni.co/> (дата звернення 20.11.2022р) – Назва з екрана.
11. Monster.com [Електронний ресурс] – Режим доступу: <https://monster.com/> (дата звернення 20.11.2022р) – Назва з екрана.
12. USAJobs [Електронний ресурс] – Режим доступу: <https://www.usajobs.gov/> (дата звернення 20.11.2022р) – Назва з екрана.

13. Glassdoor [Електронний ресурс] – Режим доступу: <https://www.glassdoor.com/index.htm/> (дата звернення 20.11.2022р) – Назва з екрана.
14. Карась К. С. Розроблення модуля «Сайт пошуку роботи». Тези доповідей
15. Чернова Т. (2021). Проектування та розробка системи пошуку кандидатів на роботу з використанням можливостей Spring Framework та Neo4j.
16. Опитування <https://dou.ua/lenta/articles/job-survey-results-2019/>
17. Faberman, R.J., Mueller, A.I., Şahin, A. and Topa, G. (2022), Job Search Behavior Among the Employed and Non-Employed. *Econometrica*, 90: 1743-1779. <https://doi.org/10.3982/ECTA18582>
18. Булелик, Я. М., Романчукевич, М. Й. (2019). Алгоритм процесу підбору персоналу в сучасних умовах.
19. Bogle, Salathiel, Sankaranarayanan, Suresh. Job Search System in Android Environment-Application of Intelligent Agents (2012). *International Journal of Information Sciences and Techniques (IJIST)* Vol.2, No.3, May 2012, Available at SSRN: <https://ssrn.com/abstract=3859843>
20. Bogle, Salathiel, Suresh Sankaranarayanan. "Intelligent Agent based Job Search System in Android Environment." 2011 IEEE International Conference on Electro/Information Technology, 2011
21. Bellifemine, Fabio, Giovanni Caire, Dominic Greenwood. *Developing multi-agent systems with Jade*. John Wiley & Sons, Ltd, 2007.
22. Bagues, M., Labini, M.S. (2007), "Do online labor market intermediaries matter? The impact of AlmaLaurea on the university-to-work transition", Working Paper No. 13621, National Bureau for Economic Research, Washington, DC.
23. Job search on the internet and its outcome. Available from: https://www.researchgate.net/publication/241703157_Job_search_on_the_internet_and_its_outcome [accessed Nov 17 2022].
24. Card D., J. Kluve, A. Weber (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association* 16 (3), 894–931.

25. Cr'epon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013). Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *The Quarterly Journal of Economics* 128 (2), 531–580.
26. Cappelli, P. (2001), "Making the most of on-line recruiting", *Harvard Business Review*, Vol. 79No. 3, pp. 139-46.
27. G'urtzgen, N., A. Diegmann, L. Pohlan, and G. J. van den Berg (2018). Do digital information technologies help unemployed job seekers find a job? Evidence from the broadband internet expansion in Germany. *IFAU Working Paper Series* 2018:21.
28. Kuhn, P. and H. Mansour (2014). Is internet job search still ineffective? *The Economic Journal* 124 (581), 1213–1233.
29. Stevenson, B. (2009). The internet and job search. In *Studies of labor market intermediation*, pp. 67–86. University of Chicago Press.
30. Thoms P., R. McMasters M., R. Roberts, and D. A. Dombkowski (1999). Resume characteristics as predictors of an invitation to interview. *Journal of Business and Psychology* 13 (3), 339–356.
31. Mary Grace G. Ventura, Rex P. Bringula Effectiveness of Online Job Recruitment System: Evidence from the University of the East. *IJCSI International Journal of Computer Science Issues*, Vol. 10, Issue 4, No 1, July 2013 www.IJCSI.org
32. Резаєв Я. О. Візуалізація статистики web-сервісу пошуку роботи для користувачів: дипломний проект бакалавра: захищений 16.06.2020 / Резаєв Ярослав Олегович; Національний Авіаційний Університет. – Київ, 2020. – 118 с. <https://er.nau.edu.ua/handle/NAU/54428>
33. Кучер М.О., Нескородєва Т.В. Аналіз факторів впливаючих на рівень щастя населення методами статистичного навчання Матеріали II всеукраїнської науково-практичної конференції «Комп'ютерні технології обробки даних» (м. Вінниця, 10 грудня 2021 р.) – Вінниця, ДонНУ імені Василя Стуса, 2021. – С. 39-42.

34. Огороднік М. О., Нескородєва Т.В. Аналіз даних про мобільні додатки APPLE IOS засобами мови R. Матеріали II всеукраїнської науково-практичної конференції «Комп'ютерні технології обробки даних» (м. Вінниця, 10 грудня 2021 р.) – Вінниця, ДонНУ імені Василя Стуса, 2021. – С. 42-44.
35. Матеріали всеукраїнської науково-практичної конференції «Комп'ютерні технології обробки даних». Вінниця, Донецький національний університет імені Василя Стуса. Кафедра інформаційних технологій. 10.12.2021 р. URL: <https://jktod.donnu.edu.ua/issue/view/411>
36. Suvankulov, F., Lau, M. C. K., Chau, F. H. C. (2012). Job search on the internet and its outcome. *Internet Research*. doi: 10.1108/10662241211235662
37. L. Azzopardi, P. Thomas, N. Craswell. Measuring the utility of search engine result pages. In *Proc. SIGIR*, pp. 605–614, 2018.
38. Weber, A. Mahringer, H. (2006), "Choice and success of job search methods", IZA Discussion Papers 1939, Institute for the Study of Labor (IZA).
39. B. Carterette, E. Kanoulas, and E. Yilmaz. Incorporating variability in user behavior into systems based evaluation. In *Proc. CIKM*, pp. 135–144, 2012.
40. Li J., Arya D., Ha-Thuc V., Sinha S. How to get them a dream job? Entity-aware features for personalized job search ranking. In *Proc. KDD*, pp. 501–510, 2016.
41. Moffat A., Zobel J. Rank-biased precision for measurement of retrieval effectiveness. *ACM Trans. Inf. Sys.*, 27(1):2.1–2.27, 2008.
42. Moffat A., Bailey P., Scholer F., P. Thomas. Incorporating user expectations and behavior into the measurement of search effectiveness. *ACM Trans. Inf. Sys.*, 35(3):24:1–24:38, 2017.
43. Werbel, J.D. (2000). Relationships among career exploration, job search intensity, and job search effectiveness in graduating college students. *Journal of Vocational Behavior*, 57(3), 379-394.
44. Wicaksono, A. F. (2019, July). Measuring job search effectiveness. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1453-1453).

45. Saha, A., Arya, D. (2017, August). Generalized mixed effect models for personalizing job search. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1129-1132).
46. White R. W., Drucker S. M. Investigating behavioral variability in web search. In Proc. WWW, pp. 21–30, 2007.
47. Wicaksono A. F., Moffat A. Empirical evidence for search effectiveness models. In Proc. CIKM, pp. 1571–1574, 2018.
48. Virga D., Rusu, A. (2018). Core self-evaluations, job search behaviour and health complaints: The mediating role of job search self-efficacy. *Career Development International*, 23, 261–273. <http://dx.doi.org/10.1108/CDI-11-2017-0208>
49. Virick M., cKee-Ryan F.M. (2018). Reemployment quality, underemployment, and career outcomes. In U.-C. Klehe & E. A. J. van Hooft (Eds.), *The Oxford handbook of job loss and job search* (pp. 359 –375). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780199764921.013.012>
50. Wanberg, C. R., Van Hooft, E. A. J., Dossinger, K., van Vianen, A. E. M., & Klehe, U.-C. (2020). How strong is my safety net? Perceived unemployment insurance generosity and implications for job search, mental health, and reemployment. *Journal of Applied Psychology*, 105, 209–229. <http://dx.doi.org/10.1037/apl0000435>
51. Wanberg, C. R., Van Hooft, E. A. J., Liu, S., Csillag, B. (2018, April). Learning to network during job search: Effects of an online training rogram. Paper presented at the 33rd Annual Meeting of the Society of Industrial and Organizational Psychology, Chicago, IL

Додаток

Текст програми

```

function (object, correlation = FALSE, symbolic.cor = FALSE,
        ...)
{
  z <- object
  p <- z$rank
  rdf <- z$df.residual
  if (p == 0) {
    r <- z$residuals
    n <- length(r)
    w <- z$weights
    if (is.null(w)) {
      rss <- sum(r^2)
    }
    else {
      rss <- sum(w * r^2)
      r <- sqrt(w) * r
    }
    resvar <- rss/rdf
    ans <- z[c("call", "terms", if (!is.null(z$weights))
"weights")]
    class(ans) <- "summary.lm"
    ans$aliases <- is.na(coef(object))
    ans$residuals <- r
    ans$df <- c(0L, n, length(ans$aliases))
    ans$coefficients <- matrix(NA_real_, 0L, 4L, dimnames =
list(NULL,
      c("Estimate", "Std. Error", "t value", "Pr(>|t|)")))
    ans$sigma <- sqrt(resvar)
    ans$r.squared <- ans$adj.r.squared <- 0
    ans$cov.unscaled <- matrix(NA_real_, 0L, 0L)
    if (correlation)
      ans$correlation <- ans$cov.unscaled
    return(ans)
  }
}

```

```

}
if (is.null(z$terms))
  stop("invalid 'lm' object: no 'terms' component")
if (!inherits(object, "lm"))
  warning("calling summary.lm(<fake-lm-object>) ...")
Qr <- qr.lm(object)
n <- NROW(Qr$qr)
if (is.na(z$df.residual) || n - p != z$df.residual)
  warning("residual degrees of freedom in object suggest
this is not an \"lm\" fit")
r <- z$residuals
f <- z$fitted.values
w <- z$weights
if (is.null(w)) {
  mss <- if (attr(z$terms, "intercept"))
    sum((f - mean(f))^2)
  else sum(f^2)
  rss <- sum(r^2)
}
else {
  mss <- if (attr(z$terms, "intercept")) {
    m <- sum(w * f/sum(w))
    sum(w * (f - m)^2)
  }
  else sum(w * f^2)
  rss <- sum(w * r^2)
  r <- sqrt(w) * r
}
resvar <- rss/rdf
if (is.finite(resvar) && resvar < (mean(f)^2 + var(c(f))) *
  1e-30)
  warning("essentially perfect fit: summary may be
unreliable")
p1 <- 1L:p
R <- chol2inv(Qr$qr[p1, p1, drop = FALSE])
se <- sqrt(diag(R) * resvar)

```

```

est <- z$coefficients[Qr$pivot[p1]]
tval <- est/se
ans <- z[c("call", "terms", if (!is.null(z$weights))
"weights")]
ans$residuals <- r
ans$coefficients <- cbind(Estimate = est, `Std. Error` = se,
`t value` = tval, `Pr(>|t|)` = 2 * pt(abs(tval), rdf,
lower.tail = FALSE))
ans$aliased <- is.na(z$coefficients)
ans$sigma <- sqrt(resvar)
ans$df <- c(p, rdf, NCOL(Qr$qr))
if (p != attr(z$terms, "intercept")) {
df.int <- if (attr(z$terms, "intercept"))
1L
else 0L
ans$r.squared <- mss/(mss + rss)
ans$adj.r.squared <- 1 - (1 - ans$r.squared) * ((n -
df.int)/rdf)
ans$fstatistic <- c(value = (mss/(p - df.int))/resvar,
numdf = p - df.int, dendf = rdf)
}
else ans$r.squared <- ans$adj.r.squared <- 0
ans$cov.unscaled <- R
dimnames(ans$cov.unscaled) <-
dimnames(ans$coefficients)[c(1,
1)]
if (correlation) {
ans$correlation <- (R * resvar)/outer(se, se)
dimnames(ans$correlation) <- dimnames(ans$cov.unscaled)
ans$symbolic.cor <- symbolic.cor
}
if (!is.null(z$na.action))
ans$na.action <- z$na.action
class(ans) <- "summary.lm"
ans
}

```