

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет
«Харківський авіаційний інститут»

Факультет інтелектуальних систем управління

Кафедра комп'ютерних наук та інформаційних технологій

Пояснювальна записка
до кваліфікаційної роботи
(тип кваліфікаційної роботи)

магістра

(освітній ступінь)

на тему «Автоматизація обробки резюме для рекрутерів за допомогою
машинного навчання»

XAI.302.366.25O.122. 9650540 ПЗ

Виконав: студент 6 курсу групи № 366
Спеціальність 122 «Комп'ютерні науки»
(код та найменування)

Освітня програма Комп'ютеризація
обробки інформації та управління
Фурманюк В. О.

(прізвище й ініціали студента)
Керівник: доц. каф. 302, к.т.н.
Лещенко Ю. О.

(прізвище й ініціали)
Рецензент: Чалий С.Ф.
(прізвище й ініціали)

Харків – 2025

**Міністерство освіти і науки України
Національний аерокосмічний університет
«Харківський авіаційний інститут»**

Факультет інтелектуальних систем управління
Кафедра комп'ютерних наук та інформаційних технологій
Рівень вищої освіти другий (магістерський)
Галузь знань 12 «Інформаційні технології»
(код та найменування)
Спеціальність 122 «Комп'ютерні науки»
Освітня програма Комп'ютеризація обробки інформації та управління

ЗАТВЕРДЖУЮ
Завідувач кафедри
Олег ФЕДОРОВИЧ

(підпис) (ім'я та прізвище)
«__» листопада 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

Фурманюку Владиславу Олександровичу

(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи Автоматизація обробки резюме для рекрутерів за допомогою машинного навчання

керівник кваліфікаційної роботи Лещенко Ю.О., к.т.н., доц. каф. 302,

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом Університету № 1765-уч. від « 14 » листопада 20 25 року

2. Термін подання кваліфікаційної роботи 08.12.2025

3. Вихідні дані до проекту (роботи) датасет для аналізу резюме

4. Зміст пояснювальної записки (перелік завдань, які потрібно розв'язати) проаналізувати сучасні підходи до автоматизації процесів рекрутингу та обробки резюме; дослідити методи обробки природної мови для вилучення ключових характеристик кандидатів із текстів резюме; розглянути існуючі архітектури моделей машинного навчання для класифікації текстових документів; сформулювати та підготувати набір даних резюме; розробити та навчити модель класифікації резюме; провести порівняльний аналіз точності різних алгоритмів машинного навчання; розробити прототип програмного модуля для автоматичної обробки резюме з можливістю інтеграції у систему управління кандидатами

5. Перелік графічного матеріалу автоматизована обробка та класифікація резюме; мета та завдання; автоматизація рекрутингу як інформаційна та

технологічна проблема; порівняння rule-based та ML-підходів у скринінгу резюме; аналіз наукових досліджень і комерційних рішень; резюме як неструктурований документ для задач машинного навчання; методи попередньої обробки та векторизації тексту; методи машинного навчання для класифікації текстових документів; дані та побудова корпусу; розробка конвєсу класифікації резюме; тонке донавчання мультимовного енкодера; калібрування імовірностей; результати та аналіз базових моделей; тонке донавчання XLM-RoBERTa; експерименти в режимі інференсу на резюме; суміщення «резюме– вакансія» (match); експерименти з резюме

6. Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада Консультанта	Підпис, дата	
		завдання видав	завдання прийняв
з основної частини	Лещенко Ю.О. доц. каф. 302		
з спеціального розділу	Лещенко Ю.О. доц. каф. 302		
з економіки та управління	Малєєва Ю. А. доц. каф. 302		

Нормоконтроль _____ Олександр ПРОХОРОВ «__» _____ 20__ р.
(підпис) (імя та прізвище)

7. Дата видачі завдання «_06_» _вересня_ 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ АВТОМАТИЗАЦІЇ РЕКРУТИНГУ ТА АНАЛІЗУ РЕЗЮМЕ	06.09 – 20.09	
2	МЕТОДИ ОБРОБКИ ПРИРОДНОЇ МОВИ ТА АЛГОРИТМИ ML ДЛЯ КЛАСИФІКАЦІЇ ТЕКСТОВИХ ДОКУМЕНТІВ	21.09 – 28.09	
3	РОЗРОБКА КОНВЄСА КЛАСИФІКАЦІЇ РЕЗЮМЕ	29.09 – 20.10	
4	ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ АЛГОРИТМІВ ОБРОБКИ РЕЗЮМЕ	21.10 – 20.11	
5	ЕКОНОМІЧНИЙ РОЗДІЛ	20.11 – 25.11	

Здобувач _____ Владислав ФУРМАНЮК
(підпис) (імя та прізвище)

Керівник кваліфікаційної роботи _____ Юлія ЛЕЩЕНКО
(підпис) (імя та прізвище)

УДК 004.8

РЕФЕРАТ

Автоматизація обробки резюме для рекрутерів за допомогою машинного навчання – Фурманюк Владислав, кваліфікаційна робота магістра, Харків, Національний аерокосмічний університет «ХАІ», кількість сторінок, враховуючи додатки 102, кількість таблиць 30, кількість рисунків 27, кількість джерел літератури 15.

КЛЮЧОВІ СЛОВА: автоматизація рекрутингу, машинне навчання, обробка резюме, обробка природної мови, класифікація, кандидат, вакансія.

ЦІЛЬ: розробка та дослідження методів автоматизованої обробки та класифікації резюме з використанням технологій машинного навчання для підвищення ефективності процесу рекрутингу.

ОБ’ЄКТ ДОСЛІДЖЕННЯ: процеси автоматизованої обробки текстових резюме та їх зіставлення з вимогами вакансій у системах підтримки рекрутингу.

ПРЕДМЕТ ДОСЛІДЖЕННЯ: методи та моделі NLP і машинного навчання для автоматизованої обробки текстових резюме у системах підтримки рекрутингу.

ЗАДАЧІ: проаналізувати сучасні підходи до автоматизації процесів рекрутингу та обробки резюме; дослідити методи обробки природної мови для вилучення ключових характеристик кандидатів із текстів резюме; розглянути існуючі архітектури моделей машинного навчання для класифікації текстових документів; сформулювати та підготувати набір даних резюме; розробити та навчити модель класифікації резюме; провести порівняльний аналіз точності різних алгоритмів машинного навчання; розробити прототип програмного модуля для автоматичної обробки резюме з можливістю інтеграції у систему управління кандидатами.

МЕТОДИ ДОСЛІДЖЕННЯ: методи обробки природної мови, лексичні (TF-ID) та семантичні (трансформери) підходи.

ОТРИМАНІ РЕЗУЛЬТАТИ: розроблено конвеєр підготовки даних та експериментів і реалізовано прототипи двох ключових функцій: багатокласова класифікація ролей та суміщення «резюме–вакансія» з топ-к ранжуванням. Використання поєднання лексичних (TF-ID) і семантичних (трансформери) підходів забезпечує високу якість і помірну вартість інференсу, а калібрування імовірностей робить пороги прийняття рішень стабільними і керованими. Практичний ефект для рекрутингу – зменшення часу первинного скринінгу, підвищення узгодженості рішень, формування шорт-листів і політики «abstain» для невпевнених випадків, що підвищує довіру користувачів.

UDC 004.8

ABSTRACT

Automating resume processing for recruiters using machine learning – Vladyslav Furmaniuk, Master Thesis, Kharkiv, National Aerospace University “KhAI”, amount of pages 102, amount of tables 30, amount of figures 27, amount of literature’s sources 15.

KEYWORDS: recruitment automation, machine learning, resume processing, natural language processing, classification, candidate, vacancy.

PURPOSE OF RESEARCH: Development and research of methods for automated resume processing and classification using machine learning technologies to enhance the efficiency of the recruitment process.

OBJECT OF RESEARCH: Processes of automated text resume processing and their matching with job requirements in recruitment support systems.

SUBJECT OF RESEARCH: NLP and machine learning methods and models for automated text resume processing in recruitment support systems.

TASKS: Analyze modern approaches to recruitment process automation and resume processing; investigate natural language processing methods for extracting key candidate characteristics from resume texts; review existing machine learning model architectures for text document classification; formulate and prepare a resume dataset; develop and train a resume classification model; conduct a comparative analysis of the accuracy of different machine learning algorithms; develop a prototype software module for automated resume processing with the possibility of integration into a candidate management system.

METHODS OF RESEARCH: Natural Language Processing methods, lexical (TF-IDF) and semantic (Transformers) approaches.

RESULTS: A data preparation and experimentation pipeline was developed, and prototypes of two key functions were implemented: multi-class role classification and 'resume-job opening' matching with top-k ranking. The use of a combination of lexical (TF-IDF) and semantic (Transformers) approaches ensures high quality and moderate inference cost, while probability calibration makes decision thresholds stable and manageable. The practical effect for recruiting includes reducing primary screening time, increasing decision consistency, forming shortlists, and implementing an 'abstain' policy for uncertain cases, which increases user trust.

ЗМІСТ

ВСТУП.....	8
1 ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ АВТОМАТИЗАЦІЇ РЕКРУТИНГУ ТА АНАЛІЗУ РЕЗЮМЕ	10
1.1 Автоматизація рекрутингу як інформаційна та технологічна проблема	10
1.2 Порівняння rule-based та ML-підходів у скринінгу резюме	15
1.3 Огляд функціональних вимог рекрутерів до систем автоматизованої обробки резюме в екосистемі ATS.....	19
1.4 Аналіз наукових досліджень і комерційних рішень у сфері автоматизованого парсингу та класифікації резюме.....	21
1.5 Формалізація резюме як неструктурованого документа для задач машинного навчання.....	25
Висновки за розділом.....	29
2 МЕТОДИ ОБРОБКИ ПРИРОДНОЇ МОВИ ТА АЛГОРИТМИ ML ДЛЯ КЛАСИФІКАЦІЇ ТЕКСТОВИХ ДОКУМЕНТІВ	31
2.1 Особливості лексики та структури резюме у задачах NLP	31
2.2 Методи попередньої обробки та векторизації тексту	33
2.3 Класичні алгоритми машинного навчання для класифікації текстових документів і їхні обмеження	35
2.4 Нейронні трансформерні архітектури для аналізу резюме та використання попередньо натренованих encoder-моделей.....	38
2.5 Метрики оцінювання якості NLP-класифікації резюме	40
Висновки за розділом.....	42
3 РОЗРОБКА КОНВЕЄРА КЛАСИФІКАЦІЇ РЕЗЮМЕ	44
3.1 Оточення та відтворюваність.....	44
3.2 Дані та побудова корпусу	45
3.3 Розбиття на train/validation/test.....	52
3.4 Базові моделі: TF-IDF → Logistic Regression та ембедінги → Logistic Regression.....	52
3.5 Тонке донавчання мультимовного енкодера.....	55
3.6 Калібрування імовірностей.....	56
Висновки за розділом.....	60
4 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ АЛГОРИТМІВ ОБРОБКИ РЕЗЮМЕ.....	61
4.1 Дизайн експериментів і протокол оцінювання	61
4.2 Результати та аналіз базових моделей.....	62
4.3 Тонке донавчання XLM-RoBERTa-base: результати й абляції.....	65
4.4 Онлайн-інференс, інтерпретація top-k та робочі пороги	67

4.4 Суміщення «резюме–вакансія» (match): регресія й бінаризація.....	71
Висновки за розділом.....	78
5 ЕКОНОМІЧНИЙ РОЗДІЛ.....	79
5.1 Мета розділу.....	79
5.2 Визначення тривалості та розрахунок собівартості виконання проекту (перший варіант)	79
5.3 Визначення тривалості та розрахунок собівартості виконання проекту (другий варіант)	84
5.4 Висновки за розділом	88
ВИСНОВКИ.....	89
ПЕРЕЛІК ПОСИЛАНЬ	90
ДОДАТОК А. Графічні матеріали	92

ВСТУП

Актуальність теми дослідження.

Ринок праці швидко цифровізується, а обсяги резюме та вакансій зростають на платформах і в корпоративних системах рекрутингу Applicant Tracking System (ATS). Первинний відбір кандидатів залишається трудомістким і суб'єктивним, особливо за умов багатомовних текстів, нестандартизованих форматів і різної якості даних. Сучасні методи NLP та машинного навчання (від лексичних до семантичних моделей) створюють передумови для автоматизації класифікації резюме, вилучення ключових характеристик та суміщення «кандидат–вакансія». Водночас потреба у прозорості, керованих порогах ухвалення рішень і перевірюваній якості (калібруванні ймовірностей, оцінюванні надійності) робить дослідження й розробку прикладних рішень у цій сфері актуальними для бізнесу та служб зайнятості.

Мета дослідження – розробка та дослідження методів автоматизованої обробки та класифікації резюме з використанням технологій машинного навчання для підвищення ефективності процесу рекрутингу.

Для досягнення мети необхідно виконати наступні **завдання**:

1. Проаналізувати сучасні підходи до автоматизації процесів рекрутингу та обробки резюме.
2. Дослідити методи обробки природної мови для вилучення ключових характеристик кандидатів із текстів резюме.
3. Розглянути існуючі архітектури моделей машинного навчання для класифікації текстових документів.
4. Сформулювати та підготувати набір даних резюме.
5. Розробити та навчити модель класифікації резюме.
6. Провести порівняльний аналіз точності різних алгоритмів машинного навчання.
7. Розробити прототип програмного модуля для автоматичної обробки резюме з можливістю інтеграції у систему управління кандидатами.

Об'єкт дослідження – процеси автоматизованої обробки текстових резюме та їх зіставлення з вимогами вакансій у системах підтримки рекрутингу.

Предмет дослідження – методи та моделі NLP і машинного навчання для автоматизованої обробки текстових резюме у системах підтримки рекрутингу.

Практичне значення отриманих результатів. Розроблено конвеєр підготовки даних та експериментів і реалізовано прототипи двох ключових функцій: багатокласова класифікація ролей та суміщення «резюме–вакансія» з топ-k ранжуванням. Використання поєднання лексичних (TF-ID) і семантичних (трансформери) підходів забезпечує високу якість і помірну вартість інференсу,

а калібрування імовірностей робить пороги прийняття рішень стабільними і керованими. Практичний ефект для рекрутингу – зменшення часу первинного скринінгу, підвищення узгодженості рішень, формування шорт-листів і політики «abstain» для невпевнених випадків, що підвищує довіру користувачів. Результати та код можуть бути безпосередньо інтегровані як мікросервіс до ATS/HRM-систем.

Структура та обсяг кваліфікаційної роботи магістра. Кваліфікаційна робота магістра складається зі вступу, 5 розділів, висновків, додатків та переліку посилань. Повний обсяг кваліфікаційної роботи магістра становить 102 сторінки, в тому числі: 27 рисунків; 30 таблиць; 1 додаток на 11 сторінках; перелік посилань з 15 найменувань на 2 сторінках.