

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 1

ЗАТВЕРДЖЕНО

Науково-методичною
радою
Державного університету
«Житомирська
політехніка»

протокол від __ _____
2024 р. №__

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ для виконання курсової роботи з навчальної дисципліни ВК4.1 «МОВИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ»

для здобувачів вищої освіти освітнього ступеня «магістр»
спеціальності 121 «Інженерія програмного забезпечення»
освітньо-професійна програма «Інженерія програмного забезпечення»
факультет інформаційно-комп'ютерних технологій
(назва факультету)
кафедра комп'ютерних наук
(назва кафедри)

Рекомендовано на
засіданні кафедри
комп'ютерних наук
(назва кафедри)

24 січня 2024 р.,
протокол № 1

Розробник: к.т.н., доцент кафедри ІПЗ Андрій МОРОЗОВ
старший викладач кафедри КН Галина МАРЧУК

Житомир
2024

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 2

ЗМІСТ

Вступ.....	3
1. Структура курсової роботи	4
2. Оформлення курсової роботи	7
3. Змістовний план основної частини курсової роботи	12
РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМАТИКИ, МЕТОДІВ ТА ЗАСОБІВ ВИРІШЕННЯ ЗАДАЧІ.....	12
1.1. Аналіз задачі, засобів та методів її вирішення.....	12
1.2 Методи і алгоритми інтелектуального аналізу даних для рішення	12
Висновки до першого розділу.....	14
РОЗДІЛ 2. ПРАКТИЧНА РЕАЛІЗАЦІЯ АЛГОРИТМІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.....	14
2.1 Характеристика джерела даних для проведення аналізу.....	14
2.2 Побудова моделі аналізу і практична реалізація	18
Висновки до другого розділу	18
РОЗДІЛ 3. АНАЛІЗ РЕЗУЛЬТАТІВ, ПІДСУМКОВІ ГРАФІКИ ТА ТАБЛИЦІ.....	18
3.1 Результати проведеного аналізу в таблицях діаграмах і графіках	18
3.2 Аналіз результатів	20
Висновки до третього розділу.....	21
ТЕМАТИКА КУРСОВИХ РОБІТ.....	22
ДОДАТКИ.....	24

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 3

Вступ

Курсовий робота - вид самостійної навчально-наукової роботи з елементами дослідження, що виконується студентами протягом семестру з метою закріплення, поглиблення і узагальнення знань, одержаних за час навчання та їх застосування до комплексного вирішення конкретного фахового завдання.

Курсовий робота з дисципліни «Мови інтелектуального аналізу даних» виконується студентами спеціальності 121 «Інженерія програмного забезпечення» на 1 курсі магістратури.

Мета виконання курсової роботи полягає в отриманні практичних навичок в галузі розробки програмних засобів, що використовують методи інтелектуального аналізу даних.

На виконання курсової роботи з навчальної дисципліни у робочій навчальній програмі передбачається 20-40 годин самостійної роботи.

Захист проводиться прилюдно перед комісією у складі 2-3 викладачів кафедри, в тому числі і керівника курсової роботи. Захист супроводжується усною доповіддю протягом 5-7 хвилин з презентацією.

Загальна оцінка включає оцінку змісту та рівня виконання пояснювальної записки, оцінку доповіді студента під час захисту, своєчасність виходу студента на захист роботи, оцінку якості та функціональності розробленого програмного забезпечення.

Захищені курсові роботи зберігаються на кафедрі протягом 5 років.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 4

1. Структура курсової роботи

При написанні курсової роботи потрібно дотримуватися таких правил:

Основна структура: *Титульний лист, Завдання, Реферат, Зміст, Перелік умовних скорочень, Вступ, Основна частина (1, 2, 3 розділи), Висновки, Література, Додатки.*

– *Титульний лист* (1 стор.) оформлюють на окремому аркуші паперу (див. Додаток А).

– *Завдання* (1 стор.) оформлюється згідно зразка (див. Додаток Б).

– *Реферат* (1 стор.) оформлюється згідно зразка (див. Додаток В).

– *Зміст* (1 стор.) містить назви всіх розділів, підрозділів і обов'язково вказуються сторінки (див. Додаток Г).

– *Перелік умовних скорочень* (1 стор.). Усі прийняті у курсовій роботі малопоширені умовні позначення, символи, скорочення і терміни пояснюють у переліку умовних скорочень.

Приклад.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

CDC	Centers for Disease Control and Prevention
ВООЗ	Всесвітня організація охорони здоров'я
МОЗ	Міністерство охорони здоров'я
ІАД	Інтелектуальні аналіз даних

– *Вступ* (1-2 стор.) має чітку структуру, до якої входять *актуальність теми, об'єкт, предмет і мета дослідження.*

Об'єкт включає в себе предмет, а не навпаки. Адже предмет

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.*- 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 5

говорить про більш вузький сектор дослідження і змушує нас конкретизувати область дослідження.

Приклад.

Об'єктом дослідження даної курсової роботи є квадратне рівняння.

Предмет дослідження – способи відбору коренів в квадратних рівняннях.

– *Основна частина курсової роботи* містить 3 розділи. **Не потрібно писати слова «Основна частина».** Основна частина починається з 1 розділу, який має певну назву. Основний розділ може містити підрозділи. Нумерація підрозділів оформлюється додаванням до номеру основного розділу номеру підрозділу. Не рекомендується використовувати багато вкладених заголовків. Рівень вкладеності не може перебільшувати 4. Назви розділів, підрозділів не можуть співпадати між собою, а також не можуть співпадати з назвою курсової роботи. Більш детально основний розділ буде розглянутий у змістовному плані основної частини.

– *Висновки* (1-2 стор.) мають відображати основні результати проведеної роботи.

– *Список використаних джерел* (20 - 25 джерел), оформлюють відповідно ДСТУ 7.1:2006 "Система стандартів з інформації, бібліотечної та видавничої справи. Бібліографічний запис. Бібліографічний опис. Загальні вимоги та правила складання" (див. Додаток Д). Публікації, що цитуються в курсовій роботі, нумерують у квадратних дужках у порядку їх появи в тексті пояснювальної записки.

За допомогою **VAK.in.ua** можна швидко та легко оформити

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№- 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 6

"Список використаних джерел". Цей портал присвячений полегшенню процедури оформлення наукових джерел відповідно до вимог атестаційної комісії (ВАК) України та проходження нормативного контролю при написанні публікацій, курсових, дипломних, дисертацій та інших наукових робіт.

Приклад.

1. Prince S. Understanding Deep Learning / Simon J.D. Prince., 2023. – 986 с.
 2. Geron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow / Aurelien Geron., 2022. – 1457 с. – (3rd ed. Edition).
 3. 500 Cities: City Boundaries [Електронний ресурс]. – 2019. – Режим доступу до ресурсу: <https://chronicdata.cdc.gov/500-Cities/500-Cities-City-Boundaries/n44h-hy2j>.
 4. Goodfellow I. Deep Learning [Електронний ресурс] / I. Goodfellow, Y. Bengio, A. Courville // MIT Press – Режим доступу до ресурсу: <https://www.deeplearningbook.org/>.
-

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 7

2. Оформлення курсової роботи

Пояснювальна записка оформлюється відповідно до вимог ДСТУ 3008-95 «Документація. Звіти в сфері науки та техніки».

– Аркуші, на яких пишеться пояснювальна записка до курсової роботи мають бути формату А4 (210 x 297 мм).

– Поля: ліворуч – 30 мм, зверху, знизу – 20 мм, праворуч – 10 мм.

– Шрифт 14 пт, Times New Roman, міжрядковий інтервал – 1,5.

– Сторінки нумерують арабськими цифрами, включаючи титульний лист і додатки, проставляючи номер праворуч зверху. Номер сторінки на титульному листі не вказують.

– Текст пояснювальної записки до курсової роботи вирівнюють по ширині, заголовки вирівнюють по центру.

– Об’єм курсової роботи – 25 - 40 сторінок.

– Текст основної частини розділяють на розділи, підрозділи, пункти і підпункти. Заголовки структурних частин пояснювальної записки (наприклад, “ЗМІСТ”, “ПЕРЕЛІК СКОРОЧЕНЬ”, “ВИСНОВКИ”) та заголовки розділів друкуються великими літерами посередині рядка. Заголовки підрозділів друкують маленькими літерами, крім першої великої, з абзацу. Кожна структурна частина пояснювальної записки повинна починатися з нової сторінки.

– *Підрозділи* повинні мати порядкову нумерацію в межах кожного розділу. Номер підрозділу складається з номера розділу і порядкового номера підрозділу, відокремлених крапкою (наприклад 1.1, 1.2).

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№: 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 8

– *Ілюстрації.* Ілюстрації (креслення, рисунки, графіки, схеми, діаграми, фотознімки) слід розміщувати безпосередньо після тексту, де вони згадуються вперше, або на наступній сторінці.

На всі ілюстрації мають бути посилання у тексті. Якщо ілюстрації створені не автором пояснювальної записки, необхідно при поданні їх дотримуватись вимог чинного законодавства про авторські права. Ілюстрації можуть мати назву, яку розміщують під ілюстрацією. Номер ілюстрації складається з номеру розділу і порядкового номеру ілюстрації, відділених крапкою.

Приклад нумерації рисунків.

Для того щоб побачити можливу статистичну залежність можна побудувати парні графіки за допомогою методу pairplot, що будує попарні залежності ознак набору даних (рис.3.2).

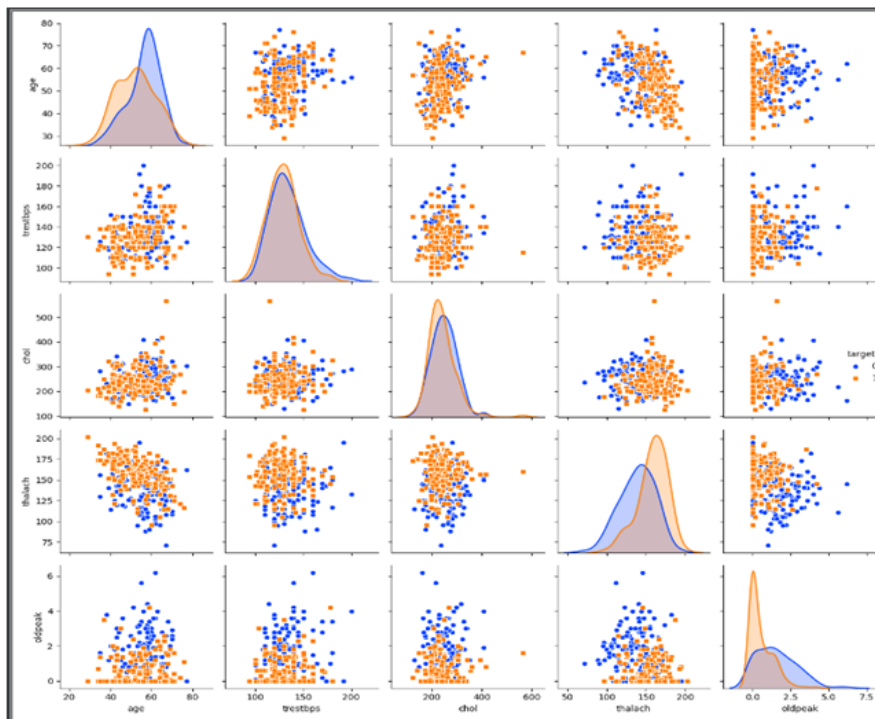


Рис. 3.2. Графіки статистичної залежності

В пояснювальній записці зазвичай дотримуються ідентичного оформлення назв рисунків (за повною або скороченою формою).

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 9

Особливо необхідно приділити увагу чіткості ілюстрацій, щоб читач без всякої напруги для очей зміг побачити окремі деталі. Особливо це стосується подання схем баз даних, діаграм, що розроблені в інших графічних редакторах і при вставці в текстовий редактор втрачають чіткість зображення.

– *Таблиці.* Таблицю слід розташовувати безпосередньо після тексту, у якому вона згадується вперше таким чином, щоб її можна було читати без повороту переплетеного блоку чи з поворотом за годинниковою стрілкою, або на наступній сторінці.

Перед заголовком таблиці та після нього, а також після таблиці пропускають по одному рядку. На всі таблиці мають бути посилання в тексті.

Номер таблиці складається з номера розділу і порядкового номера таблиці, відокремлених крапкою, наприклад, таблиця 2.1 перша таблиця другого розділу. Таблиця може мати назву, яку друкують малими літерами (крім першої великої) і вміщують над таблицею.

Слово «Таблиця» вказують один раз з права над таблицею. При перенесенні таблиці на іншу сторінку над іншими частинами пишуть: «Продовження таблиці» з зазначенням номера таблиці.

Приклад представлення таблиць.

Таблиця 1.2

Порівняння різних типів нейронних мереж

Тип нейронної мережі	Асоціативна пам'ять і розпізнавання образів	Стиснення інформації	Прогнозування	Оптимізація	Класифікація, експертні системи і діагностика	Кластеризація	Апроксимація функції	Згладжування (узагальнення)
Мережа Хопфілда	+			+				
Мережа зустрічного	+	+						

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.*- 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 10

поширення								
Мережа радіального базису (RBF)			+		+		+	+
Карта Кохонена				+	+	+		

– Для наочності подання програмного коду у записці - інтервал абзацу - «Одинарний», шрифт Courier New, напівжирний,

Приклад подання програмного коду

```
int **malloc2d(int r, int c){
    int **t=newint *[r];
    for(int i=0; i<r; i++)
        t[i]=newint[c];
    return t;
}
```

– *Формули і рівняння розташовуються безпосередньо після тексту по центру сторінки. Відстань між основним текстом і формулою становить один рядок. Номер формули ставиться на рівні формули у круглих дужках в правому положенні відносно сторінки.*

Формули набираються в редакторі формул. Формули які слідують одна за одною і які не розділені текстом розділять комою. Зручніше формулу розміщувати в невидиму таблицю.

Приклад представлення формул.

$$\frac{\delta}{\delta \theta_j} MSE(\theta) = \frac{2}{m} \sum_{i=1}^m (\theta^T * x^{(i)} - y^{(i)}) * x_j^{(i)} \quad (2)$$

– *Посилання в тексті на джерела слід зазначати порядковим номером за переліком посилань, виділеним двома квадратними дужками,*

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№- 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 11

наприклад, «у роботах [1 - 7]». Допускається наводити посилання на джерела у виносках, при цьому оформлення посилання має відповідати його бібліографічному опису за переліком посилань із зазначенням номера. При посиланнях на розділи, підрозділи, пункти, підпункти, ілюстрації, таблиці, формули, рівняння, додатки зазначають їх номери. При посиланнях слід писати: «у розділі 4», «дивись 2.1», «за 3.3.4», «на рис. 1.3», «у таблиці 3.2», «див. 3.2», «за формулою (3.1)», «у рівняннях (1.23) - (1.-25)», «у додатку Б».

– *Додатки* слід позначати послідовно великими літерами абетки, за винятком літер Г (гавва), Є, З, І, Ї, Й, О, Ч, які схожі з другими літерами або цифрами.

– Пояснювальна записка до курсової роботи здається у електронному вигляду.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№- 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 12

3. Змістовний план основної частини курсової роботи

РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМАТИКИ, МЕТОДІВ ТА ЗАСОБІВ ВИРІШЕННЯ ЗАДАЧІ

1.1. Аналіз задачі, засобів та методів її вирішення

В даному пункті описати постановку завдання на курсову роботу.

1.2 Методи і алгоритми інтелектуального аналізу даних для рішення ...

В даному пункті описати методи і алгоритми які будуть використані для рішення поставленого завдання (за тематикою курсової роботи).

Приклад.

Існує безліч алгоритмів для інтелектуального аналізу даних. Знання, що видобуваються методами Data mining, прийнято представляти у вигляді закономірностей (патернів)[9]. Серед таких виступають: асоціативні правила, дерева рішень, кластери, математичні функції.

Одним з алгоритмів машинного навчання для класифікації та прогнозування є наївний Баєсів класифікатор це простий ймовірнісний алгоритм, що базується на застосуванні теореми Баєса. Перевагами цього алгоритму є простота та його швидкість. Наївний Баєсів класифікатор відноситься до алгоритмів навчання з вчителем. Цей алгоритм часто використовують для аналізу інформації представленої в базах даних.

Враховуючи популярність та простоту алгоритму було вирішено використовувати його для рішення поставленого завдання.

1.2.1 Наївний Баєсів алгоритм

Наївний Баєсів алгоритм - це алгоритм класифікації, заснований на теоремі Баєса з припущенням про незалежність ознак. Наприклад, фрукт може вважатися яблуком, якщо він червоний, круглий і його діаметр становить близько 8 сантиметрів. Навіть якщо ці ознаки залежать одна

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 13

від одної або від інших ознак, в будь-якому випадку вони вносять незалежний внесок у ймовірність того, що цей фрукт є яблуком. У зв'язку з таким припущенням алгоритм називається «наївним».

Моделі на основі НБА досить прості і вкрай корисні при роботі з дуже великими наборами даних. При своїй простоті НБА здатний перевершити навіть деякі складні алгоритми класифікації.

Теорема Байєса дозволяє розрахувати апостеріорну ймовірність $P(c|x)$ на основі $P(c)$, $P(x)$ і $P(x|c)$ (формула 1).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

де $P(c|x)$ - апостеріорна ймовірність даного класу c (тобто даного значення цільової змінної) при даному значенні ознаки x .

$P(c)$ - апріорна ймовірність даного класу.

$P(x|c)$ - правдоподібність, тобто ймовірність даного значення ознаки при даному класі.

$P(x)$ - апріорна ймовірність даного значення ознаки.

Перевагами алгоритму є те, що класифікація, в тому числі багато класова, виконується легко і швидко. Коли допущення про незалежність виконується, НБА перевершує інші алгоритми, такі як логістична регресія (logistic regression), і при цьому вимагає менший обсяг навчальних даних. НБА краще працює з категорійними ознаками, ніж з безперервними.

Але даний алгоритм має і недоліки. Якщо в тестовому наборі даних є певне значення категорійного ознаки, яке не зустрічалося в навчальному наборі даних, тоді модель присвоїть їй нульову ймовірність і не зможе зробити прогноз. Це явище відоме під назвою «нульова частота» (zero frequency). Дану проблему можна вирішити за допомогою згладжування. Одним з найпростіших методів є згладжування Лапласа

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 14

(Laplace smoothing).

Хоча НБА є хорошим класифікатором, значення прогнозованих ймовірностей не завжди є достатньо точними. Тому не слід занадто покладатися на результати. Також перевагами цього алгоритму є його швидкість, що може бути суттєвим плюсом у випадку відсутності потужного графічного процесора.

Висновки до першого розділу

Короткі висновки про проведену роботу в межах даного розділу

РОЗДІЛ 2. ПРАКТИЧНА РЕАЛІЗАЦІЯ АЛГОРИТМІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

2.1 Характеристика джерела даних для проведення аналізу

Навести повний опис джерела даних. Якщо набір даних згенеровано особисто, потрібно описати за допомогою якого сервісу було сформовано набір даних і описати його структуру.

Приклад.

На жаль поки що медична статистика в Україні не може надати необхідного об'єму статистичної інформації для подальшого аналізу. Але можна використовувати інші джерела для проведення наукових досліджень і подальшого впровадження результатів в Українську медичну сферу. Одним з джерел цієї інформації може бути американська організація Centers for Disease Control and Prevention (скорочено CDC) - Центри з контролю та профілактики захворювань.

Для аналізу даних було обрано статистичні по хронічними хворобам в 500 містах США на основі зібраної інформації організації CDC. Ці дані самі по собі унікальні, тому що вони охоплюють 103 млн. чоловік в віці від 18 років, мають в своєму складі 27211 тисяч записів по різним територіям статистичної звітності, населення котрих складає від

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 15

50 чоловік до 26980 чоловік. Також серед даних є код штату, округу, міста, географічні координати, що дозволяє в подальшому розширити аналіз на основі інших статистичних даних як середній дохід домогосподарства, рівень безробіття та інше.

Основні цілі проєкту 500 cities:

– Цей проєкт надає дані на рівні міста та дільниць перепису, отримані з використанням методів оцінки малих територій, для 27 показників хронічного захворювання для 500 найбільших американських міст.

– Дані публікуються через публічний, інтерактивний веб-сайт «500 міст», який дозволяє користувачам переглядати, досліджувати та завантажувати дані на рівні міста та статистичної дільниці.

– Незважаючи на наявність обмежених даних на рівні округів та столиць, цей проєкт є першим у своєму роді аналізом даних, що дозволяє широко розповсюджувати інформацію для міст та малих районів у межах міст. Ця система доповнює існуючі дані спостереження, необхідні для більш повного розуміння питань охорони здоров'я, що впливають на мешканців цього міста або перепису.

– Ці високоякісні епідеміологічні дані для малих районів можуть використовуватися як окремими містами та групами міст, так і іншими зацікавленими сторонами для сприяння розробці та впровадженню ефективних та цілеспрямованих профілактичних заходів; виявлення нових проблем зі здоров'ям; встановлення та моніторинг основних цілей у сфері охорони здоров'я. Наприклад, містобудівники та виборні посадові особи можуть захотіти використовувати ці дані для орієнтації на райони з високим рівнем куріння або інших ризиків для здоров'я або ефективного втручання [5].

Показники поділяються на три основні групи:

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 16

1. Нездоровий спосіб життя (5 показників).
2. Хронічні захворювання (13 показників).
3. Охоплення населення превентивними методами (9 показників).

Усі показники представлені у вигляді відсотка населення та діапазону похибки. В Додатку А представлено опис значень показників представлених у форматі csv файлу.

Ці дані можна використовувати для різних видів статистичного аналізу. Також ці дані можна використовувати для розробки і тестування різних алгоритмів класифікації або кластеризації. Таким чином можна зробити висновок, що обрані статистичні дані підходять для проведення різних наукових досліджень.

Додаток А

Опис стовпців джерела даних

Назва поля	Опис	Тип даних
StateAbbr	Абревіатура штату	Текст
PlaceName	Назва міста	Текст
PlaceFIPS	Код міста FIPS	Текст
TractFIPS	Код зони перепису населення FIPS	Текст
Place_TractID	Комбінований код міста і зони перепису для формування унікального ідентифікатора для об'єднання з просторовими даними	Текст
Population2010	Перепис населення 2010 року	Число
ACCESS2_CrudePrev	Модельна оцінка поширеності поточної відсутності медичного страхування серед дорослих у віці 18-64 років, 2016 рік	Число
ACCESS2_Crude95CI	Розрахунковий інтервал для поширеності поточної відсутності медичного страхування серед дорослих у віці 18 - 64 років	Текст
ARTHRITIS_CrudePrev	Модельна оцінка поширеності артриту серед дорослих у віці від 18 років, 2016	Число
ARTHRITIS_Crude95CI	Розрахунковий інтервал поширеності артриту серед дорослих у віці від 18 років	Текст
BINGE_CrudePrev	Модельна оцінка поширеності споживання алкоголю серед дорослих у віці до 18 років 2016 року	Число
BINGE_Crude95CI	Розрахунковий інтервал поширеності споживання алкоголю серед дорослих у віці від 18 років	Текст
BPHIGH_CrudePrev	Модельна оцінка поширеності високого кров'яного тиску серед дорослих у віці 18 років 2015 року	Число
BPHIGH_Crude95CI	Розрахунковий інтервал поширеності високого кров'яного тиску серед дорослих у віці від 18 років	Текст
BPMED_CrudePrev	Модельна оцінка поширеності прийому ліків для контролю високого кров'яного тиску серед дорослих у віці від 18 років з високим кров'яним тиском, 2015	Число

Приклад.

В даній роботі за основу було використано файл з даними, що містить в собі категорії та фрагменти тексту заданої категорії.

Файл було завантажено онлайн за посиланням [kaggle.com](https://www.kaggle.com), що має назву `bbc-text.csv`:

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідас ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1 Арк 29 / 17	

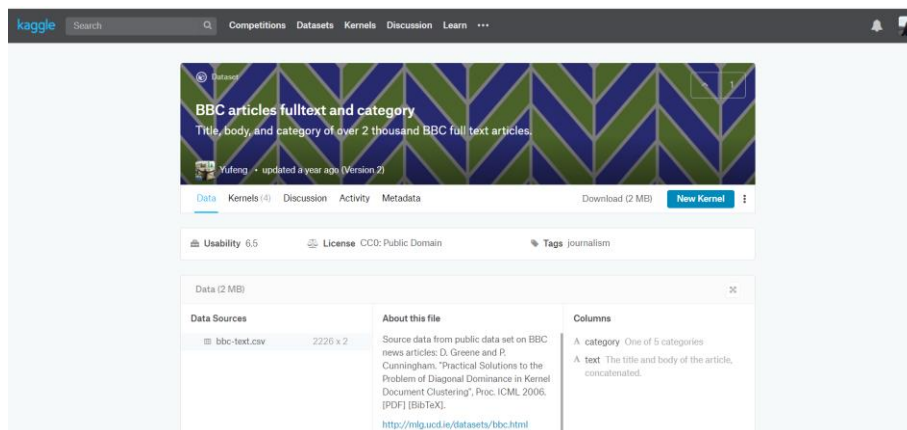


Рис.2.1. Завантаження датасету

Завантаження та огляд датасету:

```
df = pd.read_csv('C:/Users/Acer/Desktop/bbc-text.csv')
```

Виведення кількості категорій, наявних в датасеті (рис. 2.2):

```
df.groupby(by='category')['category'].count()
```

```
category
business      510
entertainment  386
politics      417
sport         511
tech          401
Name: category, dtype: int64
```

Рис.2.2. Наявні категорії та їх кількість

Наявні дані в датасеті (рис. 2.3):

```
df.head()
```

	category	text
0	tech	tv future in the hands of viewers with home th...
1	business	worldcom boss left books alone former worldc...
2	sport	tigers wary of farrell gamble leicester say ...
3	sport	yeading face newcastle in fa cup premiership s...
4	entertainment	ocean s twelve raids box office ocean s twelve...

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 18

Рис.2.3. Дані датасету

2.2 Побудова моделі аналізу і практична реалізація

Представити і описати реалізовані методи, алгоритми, використані бібліотеки.

Висновки до другого розділу

Короткі висновки про проведену роботу в межах даного розділу

РОЗДІЛ 3. АНАЛІЗ РЕЗУЛЬТАТІВ, ПІДСУМКОВІ ГРАФІКИ ТА ТАБЛИЦІ

3.1 Результати проведеного аналізу в таблицях діаграмах і графіках

Описати результати візуального аналізу.

Приклад.

Візуальний аналіз на всій вибірці проводити було складно, тому що наявно 28 стовпців даних і при попарному порівнянні виходить 784 діаграми, в кожні з яких по 27210 точок. Тому було вирішено для початку використовувати лінійну регресію для визначення кореляції показників. Результати використання лінійної регресії відображені в таблиці 3.1.

Таблиця 3.1
Результати прогнозування різних показників методом найменших квадратів

PredictBy	PredictTo	LinearRegression	PolynomialRegression	SVR
артрит	споживання алкоголю	74,90%	74,90%	74,05%
артриту	ішемічна хвороба	80,92%	81,37%	80,72%
споживання алкоголю	артрит	74,90%	75,64%	75,12%

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 19

споживання алкоголю	хронічна хвороба нирок	76,44%	76,52%	76,02%
споживання алкоголю	інсульт	82,83%	84,93%	84,13%
діагностований цукровий діабет	хронічна хвороба нирок	91,37%	91,37%	90,92%
раку (за винятком раку шкіри)	високий кров'яний тиск	0,00%	0,00%	70,19%
...

Більшість залежностей були логічними при використанні лінійної регресії. На основі лінійної регресії можна побачити наступні цікаві залежності між показниками:

1. Поширеність втрати всіх зубів після 65 років в дільницях, в яких високий показник курців. Але при подальшому аналізі було виявлено що курці рідше ходять до стоматологів (рис.3.1).

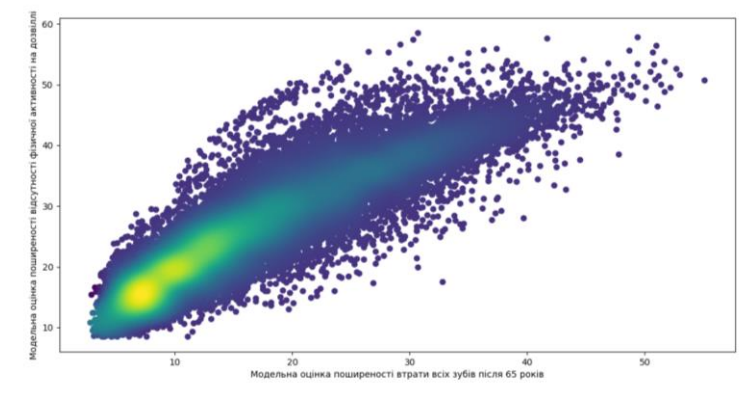


Рис.3.1. Залежність між втратою зубів та фізичною активністю

2. Наявність залежності між низькою фізичною активністю серед населення дільниці на втрату зубів. При цьому також на дільницях, де люди не займаються фізичною активністю, менша статистика візитів до стоматолога (рис.3.2).

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 20

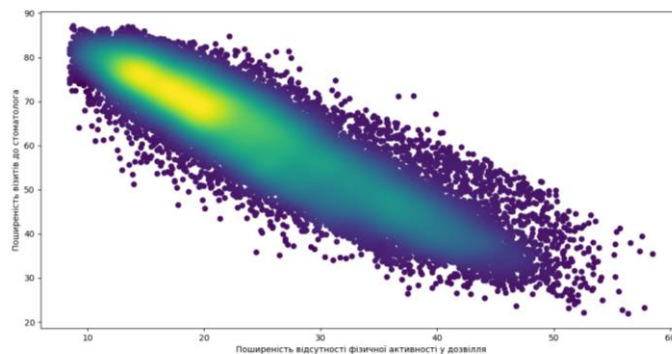


Рис.3.2. Залежність між втратою зубів та візитами до стоматолога

3.2 Аналіз результатів

За отриманими результатами зробити висновки.

Приклад.

Для аналізу була обрана статистика по хронічним захворюванням. Тестувалися не всі моделі. Серед моделей були обрані ті, які давали точність більше 70%.

На основі тестових даних можна зробити висновок, що найбільш точним методом прогнозування регресії є метод опорних векторів. Але слід зауважити що складність методу опорних векторів складає $O(n^2)$, де n – кількість записів в вибірці, що може бути недоліком у випадку наявності великої кількості навчальної вибірки.

Менш якісні результати демонструє метод поліноміальної регресії, але загалом цей вид регресії, не зважаючи на те, що залишається різновидом лінійної регресії, здатен знаходити залежності в даних, які не описуються лінійним рівнянням. При цьому використання цього методу вимагає попередньої обробки даних.

Найменш точні результати демонструє метод лінійної регресії, але цей метод регресії є найбільш простим та швидким, тому використання

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№- 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 21

лінійної регресії для розвідувального аналізу можна вважати обґрунтованим.

Таким чином можна робити висновок, що різні методи ІАД з достатньою точністю прогнозують розповсюдженість хронічних захворювань.

В результаті на деяких показниках були покращення результативності прогнозування, наприклад виявлені наступні залежності:

1. Поширеність артриту та розладів стану психічного здоров'я корелює з легеневою обструктивною хворобою на 91%, хоча по одинці ці параметри корелюють відповідно на 64.9% та 53.6%.
2. Оцінка поширеності овального аналізу крові, сигмоїдоскопії або колоноскопії та оцінка поширеності хронічної обструктивної легеневої хвороби корелюють з поширеністю кількості курців на 80,6%, хоча по одинці ці параметри корелюють відповідно на 28% та 66,5%.

Оцінка поширеності раку (за винятком раку шкіри) та оцінка поширеності відсутності фізичної активності у дозвілля корелюють з поширеністю високого рівня холестерину на 79%, хоча по одинці ці параметри корелюють відповідно на 35% та 30,8%.

Висновки до третього розділу

Короткі висновки про проведену роботу в межах даного розділу

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 22

ТЕМАТИКА КУРСОВИХ РОБІТ

1	Збір, обробка, аналіз та візуалізація вебданих (новинних потоків, Вікіпедії, блогів)
2	Розробка системи підтримки прийняття рішень для прогнозування курсу акцій
3	Розробка системи оцінки кредитоспроможності фізичних осіб з використанням методів регресійного аналізу
4	Розробка системи оцінки кредитоспроможності фізичних осіб на основі дискримінантного аналізу
5	Розробка системи розпізнавання сканованих паспортних даних
6	Розробка системи моніторингу позицій сайтів в пошукових системах
7	Розробка системи пошуку документів за допомогою k-найближчих сусідів.
8	Розробка системи прогнозування і оцінки основних економічних показників торговельного підприємства
9	Аналіз та прогнозування економічних показників за допомогою алгоритму Байєса
10	Реалізувати нейронну мережу Кохонена для вирішення завдання кластеризації даних про успішність студентів однієї з студентських навчальних груп
11	Рішення задач оптимізації за допомогою генетичних алгоритмів
12	Порівняльний аналіз нейронних мереж для розпізнавання образів
13	Розпізнавання образів на основі використання мереж Байєса
14	Застосування мережі Кохонена для класифікації особистості за психологічними ознаками
15	Провести аналіз алгоритмів глибокого навчання YOLOv3, YOLOv4 та YOLOv5
16	Провести аналіз алгоритмів глибокого навчання YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x
17	Провести аналіз алгоритмів глибокого навчання Mask R-CNN ResNet-50, Mask R-CNN ResNet-101 та Mask R-CNN FPN
18	Розробка програмного забезпечення, що моделює поведінку персонажа на основі штучного інтелекту
19	Використати інструменти Deep Reinforcement Learning і Artificial Intelligence для гри зі штучним інтелектом
20	Розробка ігрової програми «Навчання математиці» з застосуванням штучного інтелекту

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.№. 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 23

21	Аналіз текстової інформації на основі Text Mining
22	Порівняльний аналіз сучасних методів «Текстового видобутку» Text Mining
23	Розробка системи формування контрольних питань на основі матеріалів навчального курсу
24	Швидкі алгоритми аналізу формальних понять (CLOSET, FP-growth, HMiner)
25	Створення семантичної мережі для роботи з текстом
26	Класифікація текстових документів на основі технології Text Mining
27	Аналіз алгоритмів для створення систем рекомендацій
28	Розробка системи рекомендацій за допомогою алгоритму Latent Factor Model
29	Розробка системи рекомендацій за допомогою алгоритму K-Nearest Neighbors
30	Розробка системи рекомендацій на основі глибокої нейронної мережі
31	Провести аналіз методів трансферного навчання

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-* 05.01/121/**/ВК4.1 -2024
	<i>Екземпляр № 1</i>	<i>Арк 29 / 24</i>

ДОДАТКИ

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.-* 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 4

Додаток А

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»

Кафедра інженерії програмного забезпечення

КУРСОВА РОБОТА

(ПОЯСНЮВАЛЬНА ЗАПИСКА)

з дисципліни: «Мови інтелектуального аналізу даних»

на тему: «Штучний інтелект для керування неігровими
персонажами»

студента І курсу групи ПЗм-19-1
спеціальності 121 «Інженерія програмного
забезпечення»

Шевченко Ореста Івановича

(прізвище, ім'я та по-батькові)

Керівник ст. викл. каф. КН Марчук Г. В.

Дата захисту: " ___ " _____ 20__ р.

Національна шкала _____

Кількість балів: _____

Оцінка: ECTS _____

Члени комісії

(підпис)

(підпис)

(підпис)

Андрій МОРОЗОВ

(прізвище та ініціали)

Галина МАРЧУК

(прізвище та ініціали)

Дмитро ПЛЕЧІСТИЙ

(прізвище та ініціали)

Житомир – 2023

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.*- 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 4

Додаток Б

ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»

Факультет інформаційно-комп'ютерних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень: бакалавр

Спеціальність 121 «Інженерія програмного забезпечення»

«ЗАТВЕРДЖУЮ»

Зав. кафедри _____

“ ___ ” _____ 20__ р.

ЗАВДАННЯ
НА КУРСОВУ РОБОТУ СТУДЕНТУ
Шевченко Оресту Івановичу

- Тема роботи: _____,
керівник роботи: _____
- Строк подання студентом: “ ___ ” _____ 20__ р.
- Вхідні дані до роботи: _____
- Зміст розрахунково-пояснювальної записки (перелік питань, які підлягають розробці):
- _____

- _____

- _____

- Дата видачі завдання “ ___ ” _____ 20__ р.

Студент

_____ (підпис)

_____ (прізвище та ініціали)

Керівник роботи

_____ (підпис)

_____ (прізвище та ініціали)

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.-* 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 4

Додаток В

РЕФЕРАТ

Завданням на курсову роботу було створення штучного інтелекту для керування неігровими персонажами.

Пояснювальна записка до курсової роботи на тему «Штучний інтелект для керування неігровими персонажами» складається з вступу, трьох розділів, висновків, списку використаної літератури та додатку.

Текстова частина викладена на 27 сторінках друкованого тексту.

Пояснювальна записка має 27 сторінок додатків. Список використаних джерел містить 20 найменувань і займає 1 сторінку. В роботі наведено 18 рисунків. Загальний обсяг роботи – 35 сторінок.

Ключові слова: ANDROID, UNITY, ДОДАТОК, ГРА, ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ, НАВЧАЛЬНА ПРОГРАМА.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.-* 05.01/121/**/ВК4.1 -2024
	Екземпляр № 1	Арк 29 / 4

Додаток Г

ЗМІСТ

Перелік умовних скорочень	7
Вступ	8
РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМАТИКИ, МЕТОДІВ ТА ЗАСОБІВ ВИРШЕННЯ ЗАДАЧІ.....	9
1.1. Аналіз задачі, засобів та методів її вирішення.....	9
1.2 Методи і алгоритми інтелектуального аналізу даних для рішення	9
Висновки до першого розділу	11
РОЗДІЛ 2. ПРАКТИЧНА РЕАЛІЗАЦІЯ АЛГОРИТМІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.....	12
2.1 Характеристика джерела даних для проведення аналізу.....	12
2.2 Побудова моделі аналізу і практична реалізація	15
Висновки до другого розділу.....	15
РОЗДІЛ 3. АНАЛІЗ РЕЗУЛЬТАТІВ, ПІДСУМКОВІ ГРАФІКИ ТА ТАБЛИЦІ.....	16
3.1 Результати проведенного аналізу в таблицях діаграмах і графіках.....	16
3.2 Аналіз результатів.....	17
Висновки до третього розділу.....	18
Висновки	20
Список літератури.....	21
Додатки.....	22

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф.*- 05.02/2/**/**- 2024
	Екземпляр № 1	Арк 29 / 1

Додаток Д

Список використаних джерел

5. Liu G. R. Machine Learning with Python: Theory and Applications / G. R. Liu., 2022. – 692 с.
6. Mohiuddin A. Data Analytics: Concepts, Techniques, and Applications / A. Mohiuddin, K. Al-Sakib., 2020. – 450 с.
7. Practical Data Mining Techniques and Applications / K.Shah, N. Shah, V. Sawant, N. Parolia., 2023. – 214 с.
8. Bruce P. Practical Statistics for Data Scientists / P. Bruce, A. Bruce, P. Gedeck., 2020. – 350 с. – (2nd ed. edition).
9. Huyen C. Designing Machine Learning Systems / Chip Huyen., 2022. – 619 с. – (1st Edition).
10. Nelson H. Essential Math for AI / Hala Nelson., 2023. – 1035 с. – (1st Edition).
11. Chollet F. Deep Learning with Python, Second Edition / Francois Chollet., 2021. – 503 с.
12. Prince S. Understanding Deep Learning / Simon J.D. Prince., 2023. – 986 с.
13. Geron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow / Aurelien Geron., 2022. – 1457 с. – (3rd ed. Edition).
14. 500 Cities: City Boundaries [Електронний ресурс]. – 2019. – Режим доступу до ресурсу: <https://chronicdata.cdc.gov/500-Cities/500-Cities-City-Boundaries/n44h-hy2j>.
15. Goodfellow I. Deep Learning [Електронний ресурс] / I. Goodfellow, Y. Bengio, A. Courville // MIT Press – Режим доступу до ресурсу: <https://www.deeplearningbook.org/>.