Навчально-методична література

Міністерство освіти і науки України

Тернопільський національний технічний університет
імені Івана Пулюя

Кафедра комп'ютерних наук

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

Методичні вказівки до виконання курсової роботи

для здобувачів першого (бакалаврського) рівня вищої освіти всіх форм навчання зі спеціальності «Інформаційні системи і технології»

Тернопіль
2024

Козбур Г.В. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ. Методичні вказівки до виконання курсової роботи для здобувачів першого (бакалаврського) рівня вищої освіти всіх форм навчання зі спеціальності «Інформаційні системи і технології» / укладач: Козбур Г.В. – Тернопіль: ТНТУ, – 2024. – 16 с.

Укладач та відповідальний за випуск: к.т.н., доцент кафедри КН **Козбур Г. В.**

Затверджено на засіданні кафедри комп'ютерних наук, протокол № 1 від 26. 08. 2024 р.

Розглянуто та схвалено НМК факультету комп'ютерно-інформаційних систем і програмної інженерії, протокол № 1 від 02. 09. 2024 р.

Міністерство освіти і науки України

Тернопільський національний технічний університет
імені Івана Пулюя

Кафедра комп'ютерних наук

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

Методичні вказівки
до виконання курсової роботи

для здобувачів першого (бакалаврського) рівня вищої освіти
всіх форм навчання зі спеціальності
«Інформаційні системи і технології»

Тернопіль
2024

Зміст

1. Загальні відомості 5

2. Тематика курсових робіт 6

3. Основні етапи виконання КР 6

4. Структура звіту з курсової роботи 7

5. Вимоги до основних структурних частин звіту 7

6. Вимоги до оформлення звіту з курсової роботи 8

7. Оцінювання курсової роботи та академічна доброчесність 10

8. Рекомендації до презентації роботи 10

Додаток А. Рекомендації щодо самостійного формування теми курсової роботи 11

Додаток Б. Рекомендована схема виконання основної частини курсової роботи 12

Додаток В. Титульна сторінка звіту 13

1. Загальні відомості

Курсова робота – це завдання студента, яке виконується самостійно під керівництвом викладача в терміни, визначені завданням та календарним графіком їх виконання. Метою курсової роботи є закріплення, поглиблення й узагальнення знань, отриманих здобувачем вищої освіти за час навчання та їх застосування до комплексного вирішення конкретного фахового завдання.

Кожна курсова робота є індивідуальною, можлива робота в командах з чітким розмежуванням задач та відповідальності і орієнтована на розвиток у здобувачів професійних навичок і вміння творчо вирішувати практичні завдання.

Відповідно до вимог **освітньої програми «Інформаційні системи і технології»** виконання курсової роботи в рамках освоєння ОК «Інтелектуальний аналіз даних» передбачає формування та розвиток у студентів загальних та фахових **компетентностей.**

Інтегральна компетентність: Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми в області інформаційних систем та технологій, або в процесі навчання, що характеризуються комплексністю та невизначеністю умов, які потребують застосування теорій та методів інформаційних технологій.

Загальні компетентності:

КЗ01. Здатність до абстрактного мислення, аналізу та синтезу.
КЗ02. Здатність застосовувати знання у практичних ситуаціях.
КЗ05. Здатність вчитися і оволодівати сучасними знаннями.

Спеціальні (фахові) компетентності:

СК11. Здатність до аналізу, синтезу і оптимізації інформаційних систем та технологій з використанням математичних моделей і методів.

В результаті виконання та захисту курсової роботи відповідно до ОП «Комп'ютерні науки» студенти повинні досягнути таких **програмних результатів** навчання:

ПР01. Знати лінійну та векторну алгебру, диференціальне та інтегральне числення, теорію функцій багатьох змінних, теорію рядів, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації.

ПР02. Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проєктування і використання інформаційних систем та технологій.

ПР05. Аргументувати вибір програмних та технічних засобів для створення інформаційних систем та технологій на основі аналізу їх властивостей, призначення і технічних характеристик з урахуванням вимог до системи і експлуатаційних умов; мати навички налагодження та тестування програмних і технічних засобів інформаційних систем та технологій.

ПР16. Здатність інтегрувати методи інтелектуального аналізу даних у прикладні інформаційні системи для підтримки прийняття рішень, прогнозування та виявлення аномалій.

Тематика курсових робіт, **правила** та **етапи** виконання, вимоги до **оформлення** та **захисту** звітів, а також система **оцінювання** викладаються далі в «Методичних вказівках» та електронному навчальному курсі (ID: 4684 в системі електронного навчання Atutor).

2. Тематика курсових робіт

Тему курсової роботи студенти вибирають самостійно із запропонованого викладачем переліку варіантів тем або формулюють самостійно, але не пізніше визначеної дати (як правило, місяць від початку вивчення дисципліни). В цей же період студенти можуть запропонувати власні теми курсових робіт, що безпосередньо стосуються інтелектуального аналізу даних, дослідження яких для них є цікавим. У випадку, якщо студентом не вибрано теми роботи протягом зазначеного періоду, вона призначається керівником. Остаточний перелік тем курсових робіт затверджується на засіданні кафедри.

Тема затвердженої курсової роботи вноситься до **бланку завдання** встановленого зразка, який підписується здобувачем та керівником КР.

Рекомендації щодо самостійного формування тем індивідуальних завдань для КР наведені в додатку А.

3. Основні етапи виконання КР

1. Вибір напряму та теми роботи.
2. Складання плану курсової роботи.
3. Пошук і опрацювання літератури.
4. Формування завдання (завдань) курсової роботи.
5. Виконання основної частини курсової роботи.
6. Формування та оформлення звіту.
7. Презентація та захист курсової роботи.

Типові формулювання завдань, які повинен виконати студент при виконанні курсової роботи:

- виконувати попередню обробку великих масивів числових даних з допомогою інструментів електронних таблиць та мов програмування;
- розуміти принципи роботи основних методів і моделей аналізу даних;
- знати існуючі програмні продукти та різні інструментальні засоби, що використовуються в Data Mining;
- знати методи валідації побудованих моделей аналізу даних;
- читати, аналізувати та створювати візуалізації даних.

Рекомендовану схему виконання основної частини курсової роботи поміщено в додатку Б.

Пошук даних та їх підготовку до інтелектуального аналізу студент виконує самостійно.

Обсяг даних, їх доступність та відкритість, формат, опис, джерело інформації узгоджуються з керівником та надаються в роботі.

4. Структура звіту з курсової роботи

Рекомендована структура звіту з курсової роботи має такий вигляд:

- титульний лист (1 сторінка, додаток В);
- бланк завдання (1 сторінка);
- анотація (1 сторінка);
- зміст (1 сторінка);
- вступ (1-2 сторінки);
- теоретична частина (8-12 сторінок);
- практична частина (5-8 сторінок);
- висновки (1-2 сторінки);
- перелік посилань;
- додатки (лістинги, схеми, таблиці).

5. Вимоги до основних структурних частин звіту

В **анотації** дається коротка характеристика основного змісту курсової роботи та отриманих результатів дослідження. Наводиться структура роботи (курсова робота складається зі вступу, 2-х розділів, висновків, списку використаних джерел з ... найменувань; роботу викладено на ... сторінках друкованого тексту, що містить ... таблиць, ... рисунків). В кінці анотації наводяться **ключові слова** до теоретичної частини роботи (у називному відмінку, в рядок, через кому, їх загальна кількість – 3-5).

Рекомендована структура **вступу**:

- мета курсової роботи (поглиблення теоретичних знань з … та практичних навичок щодо моделей, методів та технік інтелектуального аналізу даних);
- завдання теоретичної частини, та роль в теорії інтелектуального аналізу даних та інших дисциплінах;
- завдання практичної частини курсової роботи та уточнена постановка задачі.

Рекомендовані структурні елементи **теоретичної частини**:

- місце питання в інтелектуальному аналізі даних;
- аналіз основних термінів, що входять до понятійно-категоріального апарату досліджуваної проблеми;
- клас задач, до якого підхід (метод) може бути застосований;
- суть існуючих підходів (методів), їх порівняння;
- математичний апарат підходу (методу);
- функціональна схема реалізації (алгоритм, блок-схема);
- приклади застосування підходу (методу);
- приклади реалізації з допомогою інформаційної технології, пакету прикладних програм, програмного коду, тощо).

Рекомендована структура **практичної частини:**

- повний текст практичного завдання (відповідно до індивідуального завдання);
- тип задачі та можливі варіанти вирішення для даного типу задач;
- обґрунтування методу, вибраного для вирішення завдання;
- алгоритм розв'язування задачі (з поясненнями, формулами, схемами, рисунками, таблицями, тощо);
- розв'язування з допомогою інформаційної технології чи з допомогою програмної реалізації;
- аналіз результатів, візуалізації, висновки.

Теоретична та практична частини обов'язково повинні містити посилання на літературні та інтернет-джерела інформації.

6. Вимоги до оформлення звіту з курсової роботи

6.1. Фонти і поля, якість друку

Текст курсової роботи друкується шрифтом Times New Roman, кегль 14 на одній стороні стандартного аркуша формату А4 через інтервал 1,5. Відступи перед та після абзаців – відсутні. Текст розділів вирівнюється за шириною сторінки.

Дотримуються таких розмірів полів: зліва – 2,5 см, з правого боку, знизу та зверху – 1,5 см. Усі лінії, літери, цифри і знаки повинні бути чіткими. Абзацний відступ – 1,25 см.

6.2. Розміщення структурних розділів, заголовки

Тексти анотації, змісту, вступу, кожного розділу, висновків, переліку посилань на використані літературні джерела і додатків починаються з нової сторінки.

Заголовки структурних частин курсової роботи «АНОТАЦІЯ», «ЗМІСТ», «ВСТУП», «НАЗВА РОЗДІЛУ», «ВИСНОВКИ», «ПЕРЕЛІК ПОСИЛАНЬ», «ДОДАТКИ» друкуються великими літерами та центруються.

6.3. Нумерація та розміщення структурних частин

Розділи нумеруються (1, 2), номер розділу ставлять перед назвою, НЕ зазначаючи слова РОЗДІЛ.

Підрозділи (пункти) нумерують у межах кожного розділу. Номер підрозділу складається з номера розділу і порядкового номера підрозділу, між якими ставлять крапку. В кінці номера підрозділу крапка не ставиться наприклад: «2.3» (третій підрозділ другого розділу). Потім у тому ж рядку розміщується заголовок підрозділу. Перед новим підрозділом робиться пропуск рядка. Перший абзац підрозділу повинен міститись на тій же сторінці, що і заголовок підрозділу.

Додатки нумеруються літерами (А, Б,…).

6.4. Нумерація сторінок

На титульному аркуші, бланку завдання, анотації та змістові номер сторінки не ставлять, проте враховують у загальну нумерацію (починаючи з титульного аркуша). На

наступних сторінках номер проставляють у правому нижньому куті сторінки арабськими цифрами без крапки в кінці. Сторінки додатків до загальної нумерації сторінок не відносять.

6.5. Рисунки

Рисунки (скріншоти, схеми, графіки) подаються безпосередньо після тексту, де вони згадані вперше, або на наступній сторінці.

Рисунки, що займають окремі сторінки курсової роботи, виносяться в додатки.

Рисунки нумерують послідовно в межах розділу, за винятком ілюстрацій, поданих у додатках. Номер рисунка повинен складатися з номера розділу і порядкового номера ілюстрації, між якими ставиться крапка. Наприклад: Рисунок 1.2 (другий рисунок першого розділу). Рисунок та підпис до нього центруються. Перед рисунком та після підпису рисунка робиться пропуск рядка. Між рисунком та його підписом пропуск рядка не робиться.

6.6. Формули

Формули відділяють від тексту зверху та знизу пропущеними рядками та позначають певним номером, написаним у дужках. Формули (якщо їх більше, ніж одна) нумерують у межах розділу. Номер формули складається з номера розділу і порядкового номера формули в розділі, між якими ставлять крапку. Номери формул пишуть біля правого поля аркуша на рівні відповідної формули у круглих дужках, наприклад: (3.1) (перша формула третього розділу). Посилання на формули вказують порядковим номером формули в дужках, наприклад «... у формулі (3.1)».

Пояснення значень символів і числових коефіцієнтів наводять безпосередньо під формулою.

6.7. Перелік посилань

Перелік посилань розміщується за порядком їх появи в тексті. Посилання на літературні джерела в основній частині курсової роботи наводиться в квадратних дужках. Приклади оформлення переліку посилань: https://surl.li/kzejez .

6.8. Додатки

Ілюстративний матеріал, що виноситься у **додатки**, має безпосередньо стосуватися теми дослідження. Таблиці, лістинги, графіки та схеми повинні мати назви і бути пронумеровані. **В тексті курсової роботи обов'язково повинні бути посилання на додатки** і пояснення змісту кожної таблиці, рисунка, схеми тощо. Посилання на додатки оформлюються круглими дужками (додаток В).

Оформлюючи додатки**,** на окремому аркуші друкують по центру слово «ДОДАТКИ» (Bold, кегль 48), після чого на наступних аркушах розміщують необхідні додатки, кожен додаток з нової сторінки.

Додаток повинен мати заголовок, надрукований угорі малими літерами з першої великої літери. Посередині рядка над заголовком друкується слово «Додаток __» і велика літера, що позначає додаток.

7. Оцінювання курсової роботи та академічна доброчесність

Курсова робота є окремим видом навчальної діяльності. У виконаних індивідуальних завданнях не має бути академічного плагіату, фальсифікації та фабрикації.

При оцінюванні курсової роботи беруться до уваги:

- дотримання вимог до оформлення курсових робіт;
- зміст дослідження **першого розділу** (відповідність назви теми її реальному змістові; аналіз та узагальнення проведеного огляду літератури; глибина опрацювання теоретичного матеріалу; коректність виносок і посилань);
- зміст та обсяг **другого розділу –** практичної частини курсової роботи (правильність визначення типу задачі, методу розв'язання та відповідність використання методу, інформаційної технології, мови програмування, середовища реалізації);
- захист роботи (доповідь та презентація, відповіді на запитання членів комісії).

Оцінювання курсової роботи

Відповідність курсової роботи вимогам ЄСКД	Викладення змісту першого розділу	Викладення змісту другого розділу	Наявність належно оформлених вступу, висновків, переліку посилань і додатків	Презентація та захист	Сума
5	**30**	**30**	**10**	**25**	**100**

Підсумкова оцінка за 100-бальною шкалою виставляється як усереднене значення оцінок членів комісії, виставлених окремо за виконання та захист КР.

8. Рекомендації до презентації роботи

Слайд 1. Титульний – міністерство, ВУЗ, кафедра, тема курсової роботи, прізвище та ім'я студента, група, керівники, Тернопіль, рік.

Слайд 2. Зміст роботи.

Наступні 2-3 слайди – перший розділ (стисло викладена суть теоретичного завдання, підкріплена фото, рисунками, таблицями, малюнками для наочної демонстрації).

Наступні 4-6 слайдів – третій розділ (схематичне або табличне зображення умови практичного завдання на 1 слайді, на наступних 1-2 слайдах – суть та математичний аппарат застосованого методу розв'язання, на наступних 2-3 – реалізація).

Передостанній слайд – висновки до результатів роботи.

Останній слайд – Дякую за увагу!

Слайди презентації слід занумерувати.

Оптимальні шрифти - Arial або Times New Roman (для заголовків –24 пт, для інформації –18-20 пт). Слід уникати заповнення усього слайда лише текстом. Уникати анімації при переходах між слайдами.

Загальний час презентації – 5-7 хвилин.

Додаток А

Рекомендації щодо самостійного формування теми курсової роботи

Тему курсової роботи рекомендовано формувати відповідно до схеми:

Задача – Поле – Метод – Технологія.

Наприклад,
«Сегментація клієнтів страхової компанії методом агломеративної кластеризації засобами Python»

або

«Застосування лінійної регресії до задачі оцінки вартості вживаного транспортного засобу із використанням пакету WEKA».

Індивідуальні завдання формуються на основі таблиці:

Задача	**Поле**	**Метод**	**Технологія**
• Статистичний аналіз • Класифікація • Регресія • Кластеризація (сегментація) • Асоціативні правила • Створення інформаційної панелі • Інше	• Банківська сфера • Страхування • Бізнес • Освіта • Медицина • Спорт, культура • Управління • Інше	Відповідно до задачі	• Електронні таблиці • Мова програмування Python • Мова програмування R • Середовище • Програмний продукт • Інше

Додаток Б

Рекомендована схема виконання основної частини курсової роботи

1. Вибір та/чи збір даних

- Знайти відкритий набір структурованих даних. Обсяг даних повинен дозволяти побудову моделі машинного навчання на ПК.
- Ознайомитись із назвами та змістом атрибутів даних. Здійснити опис даних (загальний обсяг в байтах, кількість рядків та стовпців, описати зміст атрибутів).
- Визначити, чи містить набір даних вектор міток. Якщо так, то визначити тип даних у векторі міток.
- Визначити, яку задачу і з допомогою якої моделі машинного навчання можна вирішити для вибраних даних.
- Завантажити дані та зберегти їх у форматі CSV.
- Сформулювати повністю тему і мету курсової роботи.

2. Підготовка даних

- Визначити тип кожного з атрибутів, число різних значень в кожному з атрибутів, усі можливі значення кожного з атрибутів або інтервал, якому вони належать. Визначити кількості пропущених значень для кожного з атрибутів.
- Вибрати стратегію для роботи з пропущеними значеннями. Виконати очистку даних (заповнити або видалити пропущені значення).
- Перетворити категоріальні змінні у числові (наприклад, використати label - encoding або one-hot encoding).
- За потреби нормалізувати або стандартизувати числові змінні.

3. Розвідувальний аналіз даних

- Провести описову статистику даних за усіма атрибутами (середнє, медіана, стандартне відхилення).
- Побудувати розподіли даних за кожним з атрибутів, виявити викиди. Вибрати стратегію для роботи з викидами.
- За потреби визначити можливі кореляції між різними атрибутами або атрибутами і міткою. Вибрати стратегію для роботи з мультиколінеарними ознаками.

4. Розробка моделі машинного навчання

- Вибрати 3-4 різних алгоритмів машинного навчання для вирішення задачі, сформульованої в темі роботи.
- Провести навчання кількох моделей та отримати базові результати.
- Вибрати стратегію оцінювання якості побудованих моделей. Визначити якість кожної з моделей.
- Запропонувати шляхи покращення якості побудованих моделей.

5. Підготовка звіту

- Оформити результати в зрозумілій формі (графіки, таблиці).
- Оформити звіт до курсової роботи. Сформулювати висновки та рекомендації на основі отриманих результатів.
- Підготувати презентацію та доповідь на 5-7 хв за результатами проведеної роботи.

Додаток В

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)
кафедра комп'ютерних наук
(повна назва кафедри)

КУРСОВА РОБОТА

з **Інтелектуального аналізу даних**
(назва дисципліни)

на тему: _____

Студента (ки) _____ курсу, групи _____
напряму підготовки _____

спеціальності _____

(прізвище та ініціали)
Керівник: к.т.н., доц. Козбур Г.В.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)
Оцінка за національною шкалою _____
Кількість балів: _____ Оцінка ECTS _____

Члени комісії: _____ _____
(підпис) (прізвище та ініціали)
_____ _____
(підпис) (прізвище та ініціали)
_____ _____
(підпис) (прізвище та ініціали)

м. Тернопіль – 202 _

Козбур Г.В. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ. Методичні вказівки до виконання курсової роботи для здобувачів першого (бакалаврського) рівня вищої освіти всіх форм навчання зі спеціальності «Інформаційні системи і технології» / укладач: Козбур Г.В. – Тернопіль: ТНТУ, – 2024. – 16 с.