

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

**Т. А. Ліхоузова,
Ю. О. Олійник**

АНАЛІЗ ДАНИХ В ІНФОРМАЦІЙНИХ СИСТЕМАХ

КУРСОВА РОБОТА

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за освітньою програмою «Інженерія програмного забезпечення інформаційних систем»
спеціальності «Інженерія програмного забезпечення»

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2022

Рецензенти: *Солдатова М. О.*, к.т.н., доцент кафедри ІСТ ФІОТ
Ткач М. М., к.т.н., доц., доцент кафедри ІСТ ФІОТ

Відповідальний
редактор *Корнага Я. І.*, д.т.н., проф., в.о. декана ФІОТ

*Гриф надано Методичною радою КПП ім. Ігоря Сікорського
(протокол № 1 від 02.09.2022 р.)
за поданням Вченої ради факультету інформатики та обчислювальної техніки
(протокол № 11 від 11.07.2022 р.)*

Посібник призначений для набуття практичних навичок аналізу даних та побудови моделей при вивченні студентами дисципліни «Аналіз даних в інформаційних системах». Обсяг та перелік тем запропонованого практичного матеріалу повністю покриває потреби 121 спеціальності.

Реєстр. № НП 22/23-032. Обсяг 0,8 авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
<https://kpi.ua>

Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

© Т. А. Ліхоузова, Ю. О. Олійник, 2022

© КПП ім. Ігоря Сікорського, 2022

ЗМІСТ

ЗАГАЛЬНІ ПОЛОЖЕННЯ	4
ПОРЯДОК ВИКОНАННЯ КУРСОВОЇ РОБОТИ	6
Загальна характеристика предметної області та постановка задачі	6
Розробка моделі сховища даних (опціонально)	6
Розробка ETL процесів (опціонально)	6
Завантаження даних за допомогою ETL процесів (опціонально)	6
Обґрунтування вибору методів та моделей інтелектуального аналізу даних	7
Застосування та порівняння ефективності методів та моделей інтелектуального аналізу даними для поставленого завдання	7
Оформлення пояснювальної записки	7
Захист курсової роботи	7
ЗМІСТ ПОЯСНЮВАЛЬНОЇ ЗАПИСКИ	8
ПРАВИЛА ОФОРМЛЕННЯ ПОЯСНЮВАЛЬНОЇ ЗАПИСКИ	12
Форматування	13
Нумерація	13
Ілюстрації	14
Таблиці	15
Формули	15
Оформлення цитат і переліку посилань	15
Оформлення додатків	15
ПОРЯДОК ПЕРЕВІРКИ КУРСОВОЇ РОБОТИ	17
Хід виконання та захисту КР	17
Критерії оцінювання КР	17
Календарний план	19
ТЕМИ ІНДИВІДУАЛЬНИХ ЗАВДАНЬ	20
СПИСОК ЛІТЕРАТУРИ	21
ДОДАТКИ	22

1 ЗАГАЛЬНІ ПОЛОЖЕННЯ

Виконання курсової роботи (КР) з дисципліни «Аналіз даних в інформаційних системах» є обов'язковою складовою навчального плану підготовки бакалаврів за спеціальністю 121 «Інженерія програмного забезпечення».

Головна мета КР полягає у закріпленні, поглибленні та узагальненні базових теоретичних знань, якими студент оволодів під час вивчення дисципліни «Аналіз даних в інформаційних системах», їх застосуванні до комплексного вирішення конкретного фахового завдання. Вона охоплює всі аспекти інженерії аналізу даних та побудов сховищ даних.

Під час вивчення дисципліни студент отримує відповідну теоретичну та практичну підготовку, яка сприяє розширенню наукового кругозору майбутнього спеціаліста, забезпечує підвищення продуктивності праці за рахунок ефективного використання сучасних технічних та технологічних засобів, дозволяє йому успішно опановувати суміжні спеціальності.

Дана дисципліна складається з одного кредитного модуля «Курсова робота з аналізу даних в інформаційних системах». Обсяг дисципліни – 1 кредит ECTS.

Для успішного виконання курсової роботи студентам необхідно мати базові знання із таких дисциплін:

- «Алгоритми та структури даних»;
- «Об'єктно - орієнтоване програмування»;
- «Бази даних»;
- «Аналіз даних в інформаційних системах».

Успішне оволодіння знаннями з даної дисципліни готує студентів до дипломного проектування.

Основними цілями написання курсової роботи є:

- вивчення сучасних концепцій та підходів до оперативного аналізу даних та створення сховищ даних;

- опанування практичних прийомів з виконання аналізу даних;
- отримання навичок побудови та верифікації моделей.

Робота виконується студентами усіх форм навчання протягом четвертого семестру.

Під час виконання курсової роботи студент повинен продемонструвати:

знання:

- фундаментальних основ, термінів та понять в області сховищ даних та оперативного аналізу даних;
- загальних принципів проведення аналізу даних;
- методів побудови та верифікації моделей на основі статистичних даних;

уміння:

- застосовувати архітектури сховищ відповідно до класу задач;
- застосовувати методи ETL;
- моделювати багатовимірні моделі сховищ даних та застосовувати CASE засоби для моделювання;
- застосовувати відповідний інструментарій для візуалізації даних;
- застосовувати відповідний інструментарій для проведення аналізу даних;
- застосовувати різні статистичні моделі для опису поведінки досліджуваних об'єктів;

досвід:

- моделювання структуру сховища даних;
- розробки рішення з аналізу даних;
- вибору, побудови та верифікації статистичних моделей.

2 ПОРЯДОК ВИКОНАННЯ КУРСОВОЇ РОБОТИ

Основні етапи виконання курсової роботи:

- постановка задачі;
- розробка сховища даних (опціонально)
 - а) розробка моделі сховища даних;
 - б) розробка ETL процесів;
 - в) завантаження даних за допомогою ETL процесів;
- обґрунтування вибору методів інтелектуального аналізу даних та критеріїв оцінки якості моделей;
- формування навчальних та тестових вибірок;
- застосування та порівняння ефективності методів інтелектуального аналізу даних для поставленого завдання;
- оформлення пояснювальної записки;
- підготовка презентації та захист курсової роботи.

2.1 Загальна характеристика предметної області та постановка задачі

Постановка задачі є самостійним етапом роботи по виконанню КР. На цьому етапі визначається предметна область, робиться огляд доступних джерел даних і наводиться опис постановки завдання аналізу даних.

2.2 Розробка моделі сховища даних (опціонально)

На цьому етапі виконується опис джерел даних та розробляється модель сховища даних у CASE засобі (наприклад Power Designer), формати даних.

2.3 Розробка ETL процесів (опціонально)

На цьому етапі розробляються ETL процеси для завантаження даних з зовнішніх джерел. Зовнішні джерела можуть бути представлені в вигляді окремих фалів, баз даних, зовнішніх систем з API інтерфейсом

2.4 Завантаження даних за допомогою ETL процесів (опціонально)

На цьому етапі відбувається первинне завантаження даних та налаштування періодичного завантаження даних.

2.5 Обґрунтування вибору методів та моделей інтелектуального аналізу даних

На цьому етапі відбувається теоретичне обґрунтування вибору методів інтелектуального аналізу даних для поставленої задачі. Необхідно провести аналіз літератури з використанням авторитетних наукових джерел (матеріали міжнародних конференцій, сайти arxiv.org і т.п.).

2.6 Застосування та порівняння ефективності методів та моделей інтелектуального аналізу даними для поставленого завдання

На цьому етапі відбувається застосування та порівняння методів інтелектуального аналізу даних в рамках поставленого завдання. Необхідно застосувати не менше 3-х методів (наприклад класифікація та кластеризація) та порівняти ефективність.

2.7 Оформлення пояснювальної записки

Пояснювальна записка оформлюється згідно вимог викладених у третьому та четвертому розділах даних методичних вказівок.

2.8 Захист курсової роботи

Захист курсової роботи передбачає демонстрацію студентом результатів роботи у формі презентації. Студент має показати володіння матеріалом курсової роботи. Повністю завершена програма та оформлена пояснювальна записка повинні бути надані за тиждень до дати захисту.

3 ЗМІСТ ПОЯСНЮВАЛЬНОЇ ЗАПИСКИ

Обов'язковими структурними частинами пояснювальної записки є:

- титульний лист;
- лист завдання на курсову роботу;
- анотація;
- зміст;
- вступ;
- основна частина;
- висновок;
- перелік посилань;
- додатки до пояснювальної записки.

Титульний лист повинен бути встановленого зразку. На ньому вказується назва міністерства, університету, інституту, факультету, кафедри і тема курсової роботи (у точній відповідності із індивідуальним завданням). Його зразок наведений у *Додатку Б*. Титульний лист не нумерується як розділ, не вноситься до змісту і не нумерується як сторінка.

Лист завдання на курсову роботу видається і заповнюється керівником. Заповнення листа завдання здійснюється викладачем при кожній консультації та здачі КР. Бланк завдання на курсову роботу повинен бути оформлений за формою, наведеною в *Додатку В*. На ньому обов'язково повинна бути вказана дата видачі завдання. Індивідуальне завдання засвідчується підписом керівника КР. Лист завдання не нумерується як розділ.

В **анотації** у реферативному стилі наводиться інформація про зміст та результати, що отримані в курсовій роботі. Як розділ анотація не нумерується.

Зміст характеризує структуру КР. Він повинен вміщувати в собі назви усіх розділів, підрозділів, пунктів та підпунктів курсової роботи, а також перелік додатків. Усі назви повинні бути записані так само як вони сформульовані в КР. Визначення сторінок у змісті обов'язкове. Зміст розміщується на окремій

сторінці, як розділ зміст не нумерується. Посилання на зміст, у сам зміст не включається.

У **вступі** коротко розкривається призначення КР, ціль роботи, сутність вирішуваної задачі, дається загальна постановка завдання і методи його розв'язання. Вступ як розділ не нумерується.

Основна частина пояснювальної записки може складатися з розділів, підрозділів, пунктів, підпунктів^{1*}. Кожний розділ починають з нової сторінки.

В *першому розділі* висвітлюється інформаційна сутність задачі – описується предметна область, наводиться опис джерел даних. В розділі треба конкретизувати, кому і навіщо потрібне це дослідження, що буде вхідними даними і в якому вигляді треба надати результат. Типові підрозділи:

- Огляд предметної області — короткий огляд предметної області, виявлення проблем, що потребують для вирішення поглибленого аналізу даних; вибір проблеми, для якої буде виконано аналіз; 1-3 сторінки.
- Огляд доступних джерел даних — огляд інформації про те, що вже зроблено іншими дослідниками для вирішення проблеми; збір інформації про доступні джерела даних і вибір переліку джерел, що будуть використані в роботі.
- Постановка задачі — формулювання мети дослідження та уточнення питань, на які повинен дати відповідь проведений аналіз.

У другому та подальших розділах розкриваються задачі згідно календарного плану курсової роботи. Студент має продемонструвати свою обізнаність в питаннях теорії і методології тих методів, які використовуються для розв'язання поставленої задачі.

У *другому розділі* розкриваються процеси збору та попередньої обробки даних (цей розділ може бути пропущений, якщо знайдено джерело даних, для яких попередня обробка не потрібна). Типові підрозділи:

- Розробка моделі сховища даних — наводиться опис джерел даних та

^{1*} Дотримання в курсовій роботі запропонованої рубрикації не є обов'язковою вимогою.

розробляється модель сховища даних у CASE засобі (наприклад Power Designer), обираються формати даних.

- Розробка ETL процесів — наводиться опис ETL процесів для завантаження даних з зовнішніх джерел. Зовнішні джерела можуть бути представлені в вигляді окремих фалів, баз даних, зовнішніх систем з API інтерфейсом.

- Збір та попередня підготовка даних для аналізу — наводиться описова статистика для зібраних даних, можна навести графіки. Якщо потрібна очистка даних, то вказати для чого і як саме вона виконується.

У *третьому розділі* описують застосування методів та моделей інтелектуального аналізу даних та виконують їх порівняння ефективності для поставленого завдання. Типові підрозділи:

- Визначення моделей та методів, що можуть бути використані — робиться теоретичне обґрунтування вибору методів інтелектуального аналізу даних для поставленої задачі. Необхідно провести аналіз літератури з використанням авторитетних наукових джерел (матеріали міжнародних конференцій, сайти arxiv.org і т.п.).

- Вибір ознак, що будуть використані для аналізу — якщо використовуються не всі доступні ознаки, то їх вибір обґрунтовується.

- Підготовка даних для навчання та верифікації моделей — вказати, що саме було виконано для підготовки даних (способи обробки невизначених значень та викидів, нормалізація, тощо; як саме дані були розділені на навчаючу та тестову вибірки; які критерії буде використано для оцінки якості моделей).

- Формування моделей — вибір оптимального класу складності моделей (вказати, які критерії використано для вибору класу складності для кожного типу моделі, навести графіки).

- Верифікація моделей — навести результати верифікації моделей на тестових даних, бажано порівняти на графіку різні моделі.

- Висновки щодо якості побудованих моделей — вказати точність кожної моделі, яка з них краща на даному етапі, яку з них можна легше скоректувати при зміні/збільшенні навчальної вибірки.
- Результати аналізу — відповіді на питання, що були поставлені перед дослідженням.

У **висновках** в реферативній формі повинні бути описані результати, отримані студентом на кожному із етапів виконання роботи (аналітичному, етапі проектування програмного забезпечення, експериментальному дослідженню, аналізу отриманих результатів), а також висновки щодо досягнення мети курсової роботи. Тут необхідно наголосити на якісних і кількісних показниках здобутих результатів, обґрунтувати їх достовірність. Висновки як розділ не нумеруються.

Перелік посилань повинен включати усі літературні джерела, на які є посилання у тексті пояснювальної записки. Список повинен формуватися в порядку посилань за текстом і вміщувати бібліографічні відомості офіційно виданих книжок, статей, патентів, депонованих рукописів тощо. Як розділ перелік літератури не нумерується.

В **додатки** включають опис моделі даних, лістинг скриптів ETL процесів, програмний код застосованих методів та засобів, скріншоти тощо, які не увійшли до пояснювальної записки, але потрібні для пояснень.

У *додатку А* повинні міститися лістинги/скрипти. Тексти програм/скриптів повинні мати коментарі. Текст програми (функцій) повинен відповідати розробленим алгоритмам.

У *Додаток Б* за необхідності виносять проміжні математичні доведення, формули та розрахунки; таблиці, графіки, скріншоти тощо, які не увійшли до пояснювальної записки. Як розділ додатки не нумеруються, але кожен з додатків нумерується великими літерами алфавіту згідно ДСТУ 3008-95.

4 ПРАВИЛА ОФОРМЛЕННЯ ПОЯСНЮВАЛЬНОЇ ЗАПИСКИ

Основними документами, що представляють КР, є пояснювальна записка та папка з розробленими матеріалами:

- модель сховища даних;
- ETL процеси, завантажені дані у вигляді скриптів;
- програмний код застосованих методів аналізу даних, порівняння ефективності.

Текст пояснювальної записки до курсової роботи повинен бути викладений лаконічно, у обґрунтованому стилі. Не дозволяється переписування літературних джерел та використання не опрацьованих студентом інтернет-оглядів.

Пояснювальна записка виконується на аркушах формату А4 згідно [ДСТУ 3008-95](#). У випадку необхідності окремі ілюстрації можуть виконуватись на аркушах більших форматів.

Пояснювальна записка має бути представлена в електронному та друкованому вигляді (на час дистанційного навчання обов'язковою є тільки електронна версія).

Електронна версія зберігається в банку даних кафедри. Файл із копією курсової роботи здається на кафедру разом із друкованим примірником безпосередньо під час захисту. Формат файлу – doc, docx ,pdf.

Текстову частину роботи необхідно друкувати на одному боці аркуша білого паперу формату А4 (210×297 мм). В окремих випадках, для більш наочного подання таблиць та ілюстрацій, можна використовувати папір формату А3 (297×420 мм).

На останній сторінці пояснювальної записки необхідно наклеїти конверт, в який вкласти компакт-диск. Студенти, які мають непарний варіант, наклеюють конверт у верхній частині листа, а інші - у нижній. Це потрібно для компактного розташування пояснювальних записок у архіві.

Оптимальний обсяг **основної частини** роботи (без додатків) має складати 15-25 аркушів. Обсяг додатків жорстко не лімітується, але пропонується максимальний об'єм у кількості 20 аркушів.

4.1 Форматування

Форматування пояснювальної записки:

- поля: ліве, верхнє і нижнє не менше 20мм, праве не менше 10мм;
- шрифт: Times New Roman, 14 pt;
- міжрядковий інтервал – 1,5 pt;
- відступ першого рядка – 1,25 см;
- вирівнювання: назв розділів – по центру, назв підрозділів, пунктів і підпунктів – по ширині, основного тексту – по ширині.

Заголовки структурних частин курсової роботи: «ЗМІСТ», «ВСТУП», «РОЗДІЛ», «ВИСНОВКИ», «ДОДАТКИ», «ПЕРЕЛІК ПОСИЛАНЬ» та назви розділів друкуються великими літерами. Кожну структурну частину роботи потрібно починати з нової сторінки.

4.2 Нумерація

Нумерацію сторінок, розділів, підрозділів, пунктів, підпунктів, рисунків, таблиць, формул подають арабськими цифрами без знака «№».

Першою сторінкою курсової роботи є титульний аркуш, який включають до загальної нумерації, але номер сторінки на ньому не ставлять. На всіх наступних сторінках номер проставляють у правому верхньому куті сторінки без крапки в кінці.

Такі структурні частини роботи, як «ЗМІСТ», «ВСТУП», «ВИСНОВКИ», «ДОДАТКИ», «ПЕРЕЛІК ПОСИЛАНЬ» не мають порядкового номеру.

Підрозділи 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. Потім у тому ж рядку наводять назву підрозділу.

Аналогічним чином нумерують пункти (порядкова нумерація в межах підрозділу) і підпункти (порядкова нумерація в межах пункту). Наприклад, «2.3.1» – перший пункт третього підрозділу другого розділу.

Переліки, за потреби, можуть бути наведені всередині пунктів або підпунктів. Перед переліком ставлять двокрапку, або крапку в залежності від контексту.

Перед кожною позицією переліку слід ставити малу літеру української абетки з дужкою, або, не нумеруючи – дефіс (перший рівень деталізації).

Для подальшої деталізації переліку слід використовувати арабські цифри з дужкою (другий рівень деталізації).

Приклад:

а) форма і розмір клітин;

б) живий склад клітин:

1) частини клітин;

2) неживі включення протопластів;

в) утворення тканини.

Переліки першого рівня деталізації друкують з абзацного відступу, другого рівня – з відступом відносно місця розташування переліків першого рівня.

4.3 Ілюстрації

Ілюстрації (фотографії, скріншоти, креслення, схеми, графіки, рисунки, карти) і таблиці необхідно подавати безпосередньо після тексту, де вони згадані вперше, або на наступній сторінці. Ілюстрації позначають словом «Рисунок» і нумерують послідовно в межах розділу, за винятком ілюстрацій, поданих у додатках. Номер ілюстрації повинен складатися з номера розділу і порядкового номера ілюстрації, між якими ставиться крапка. Номер, назва і пояснювальний підпис (у разі необхідності) повинні міститися безпосередньо під ілюстрацією.

4.4 Таблиці

Таблиці нумерують послідовно в межах розділу (за винятком таблиць, поданих у додатках). В лівому верхньому куті над відповідним заголовком таблиці розміщують напис «Таблиця» із зазначенням її номера. Номер таблиці повинен складатися з номера розділу і порядкового номера таблиці, між якими ставиться крапка. Якщо таблиця займає більше одного аркуша, її заголовок переноситься на наступний, в лівому верхньому куті над відповідним заголовком таблиці розміщують напис «Продовження таблиці» із зазначенням її номера, без назви таблиці.

4.5 Формули

Формули в курсовій роботі нумерують в межах розділу. Номер формули повинен складатися з номера розділу і порядкового номера формули, між якими ставиться крапка. Номер формули пишуть в круглих дужках і розміщують біля правого поля аркуша на рівні відповідної формули.

Посилання в тексті роботи на ілюстрації, таблиці, формули вказують порядковим номером в круглих дужках, наприклад, «... у формулі (2.1)».

4.6 Оформлення цитат і переліку посилань

Бібліографічний опис літературних джерел можна розміщувати одним із таких способів: у порядку появи посилань у тексті, в алфавітному порядку прізвищ перших авторів або заголовків, у хронологічному порядку.

Посилання в тексті роботи на літературні джерела вказують порядковим номером в квадратних дужках, наприклад, «... в роботі [4] показано ...». Якщо в курсовій роботі наводиться цитата, то обов'язковим є зазначення сторінок першоджерела, що містять цитовану інформацію, наприклад, «...[4, с. 10-12]». Цитати беруться в лапки і виділяються курсивом. Приклади бібліографічного опису окремих видів літературних джерел наведених в додатку Д.

4.7 Оформлення додатків

Додаток повинен мати заголовок, надрукований на початку сторінки із вирівнюванням по центру сторінки. На одну строку вище від заголовку

друкується слово «Додаток ____» і велика літера, що позначає додаток. Додатки слід позначати послідовно великими літерами української абетки, за винятком літер Г, Ґ, Є, І, Ї, Й, О, Ч, Ї.

Текст кожного додатку за необхідності може бути поділений на розділи й підрозділи, які нумерують в межах кожного додатка. В цьому разі перед кожним номером ставлять позначення додатка (літеру) і крапку, наприклад А.2.1 – перший підрозділ другого розділу додатка А. Ілюстрації, таблиці і формули нумерують аналогічним чином.

5 ПОРЯДОК ПЕРЕВІРКИ КУРСОВОЇ РОБОТИ

5.1 Хід виконання та захисту КР

Керівник здійснює контроль за ходом виконання студентом курсової роботи, надає йому необхідну консультативну допомогу. Протягом семестру студент демонструє викладачу поточні результати роботи над проектом.

Після перевірки роботи викладач призначає день, час і місце захисту. Напередодні захисту студенту необхідно повторити теоретичний матеріал, що стосується роботи, та переглянути безпосередньо її зміст.

Захист КР проводиться у формі співбесіди зі з'ясуванням усіх питань, що виникли у керівника під час перевірки роботи. За результатами захисту, у відповідності до критеріїв оцінювання, що наведені у підрозділі 5.2 даних вказівок, викладач виставляє студенту оцінку. На оцінку за КР впливають:

- якість розробленого програмного забезпечення;
- якість розробленої програмної документації;
- компетентність та загальна ерудиція студента при відповідях на запитання під час захисту;
- ступінь виконання графіку підготовки курсової роботи.

Якщо студент подав на захист не самостійно виконану роботу, про що свідчить його некомпетентність у рішеннях та матеріалах роботи, КР до захисту не допускається, що супроводжується записом "не допущений" у екзаменаційній відомості. Такий самий запис робиться у випадку, якщо КР не завершена на час захисту. В цих випадках запис "не допущений" еквівалентний отриманню оцінки "незадовільно".

5.2 Критерії оцінювання КР

Рейтингова оцінка з курсової роботи має дві складові: виконання курсової роботи та її захист. Перша (стартова) складова характеризує роботу студента з курсового проектування та її результат – якість пояснювальної записки; друга складова характеризує якість захисту студентом курсової роботи.

Розмір шкали першої складової дорівнює 80 балів, а другої – 20 балів.

Стартова складова виконання курсової роботи (r1):

- ступінь розкриття теоретичних аспектів теми та коректність використання понятійного апарату – до 5 балів;
- повнота та коректність предметної області – до 15 балів;
- обґрунтування обраних методів – до 15 балів;
- застосування та дослідження ефективності обраних методів – до 35 балів;
- якість оформлення пояснювальної записки з урахуванням виконання вимог нормативних документів – до 10 балів;

Складова захисту курсової роботи (r2):

- ступінь володіння теоретичним матеріалом – до 5 балів;
- презентація роботи та відповіді на запитання – до 15 балів.

Таблиця 5.1 – Загальний критерій оцінювання

R = r1+ r2	Оцінка
95... 100	відмінно
85 ... 94	дуже добре
75 ... 84	добре
65 ... 74	задовільно
60 ... 64	достатньо
Менше 60	незадовільно
Курсову роботу не завершено або не надано до захисту	не допущено

5.3 Календарний план

Таблиця 5.2 – Календарний план

№ п/п	Назва етапів курсової роботи	Термін виконання
1.	Ознайомлення з предметною областю, в якій буде проводитись аналіз. Виявлення проблем в предметній області, що потребують для вирішення поглибленого аналізу даних. Вибір проблеми, для якої буде виконано аналіз. Огляд доступних джерел даних	
2.	Затвердження теми курсової роботи	10 березня
3.	Збір та попередня підготовка даних для аналізу. Формування описової статистики для зібраних даних. Постановка мети дослідження та уточнення питань, на які повинен дати відповідь проведений аналіз	20 березня
4.	Визначення моделей, що можуть бути використані. Вибір ознак, що будуть використані для аналізу	30 березня
5.	Підготовка даних для навчання та верифікації моделей. Формування моделей. Вибір оптимального класу складності моделей. Верифікація моделей	15 квітня
6.	Порівняльний аналіз моделей. Висновки про можливість використання результатів моделювання для проблеми, що досліджується (прогнозування, класифікації тощо). Формування обґрунтованих відповідей на питання, що були поставлені перед дослідженням	1 травня

7.	Оформлення пояснювальної записки та презентації. Здача закінченої роботи	15 травня
8.	Демонстрація результатів аналізу у вигляді презентації	19-26 травня

6 ТЕМИ ІНДИВІДУАЛЬНИХ ЗАВДАНЬ

Теми завдань обираються студентами самостійно і погоджуються з викладачем. Бажано обирати ті теми досліджень, що будуть корисні комусь конкретному або цікаві студенту. Приклади тем:

- Дослідження моделей регресійного аналізу на прикладі даних про чистий прибуток підприємств за видами економічної діяльності.
- Дослідження методів кластеризації на прикладі даних про виїзд громадян України за кордон.
- Дослідження методів прогнозування на прикладі даних про чисельність населення в Україні.

Теми з аналізу текстової інформації:

- Аналіз української класичної художньої літератури. Комплексна тема з задіянням групи студентів.
- Виявлення плагіату в документах. Комплексна тема з задіянням групи студентів.
- Виявлення фактів розповсюдження дезінформації в новинах. Комплексна тема з задіянням групи студентів.

СПИСОК ЛІТЕРАТУРИ

- 1) Керівництво користувача PD. Електронний ресурс режим доступу:
<http://www.sybase.com/products/modelingdevelopment/powerdesigner>
- 2) Гектор Гарсиа-Молина, Джеффри Д. Ульман, Дженнифер Уидом. Системи баз даних. Повний курс, 2004.1088с. ил. ISBN 5-8459-0384-X; 2004 г.
- 3) Барсегян А.А., Куприянов М.С. Степаненко В.В., Холод И.И. Методи и модели анали за даннях: OLAP и Data Mining.- СПб.: БХВ-Петербург, 2004. 336 с.: ил. ISBN 5-94157-522-X.
- 4) W. H. Inmon, Building the Data Warehouse, Fourth Edition: Wiley Publishing, Inc. ISBN-13: 978-0-7645-9944-6, ISBN-10: 0-7645-9944-5
- 5) Jaspersoft OLAP Ultimate Guide. Навчальний посібник
- 6) Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker. The Data Warehouse Lifecycle Toolkit, 2nd Edition. Wiley Publishing, Inc, 2008.672с.:ил. ISBN: ISBN 978-0-470-14977-5.
- 7) Ralph Kimball, Margy Ross, Warren Thornthwaite (Contributions by), Joy Mundy (Contributions by), Bob Becker (Contributions by). Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Wiley Publishing, Inc, 2010. 744 .:ил. ISBN: 978-0-470-56310-6.
- 8) Пасічник В.В., Шаховська Н.Б. Сховища даних: Навчальний посібник. - Львів: “Магнолія 2006”, 2008.-496 С. ISBN 978-966-2025-18-7.

ДОДАТКИ

ДОДАТОК А Тексти програмного коду

Тексти програмного коду програмного забезпечення

<Тема курсової роботи>

(Найменування програми (документа))

CD-RW

(Вид носія даних)

арк, Кб

(Обсяг програми (документа), арк., Кб)

студента групи ІІІ-ХХ 2 курсу

<ПІБ>

ДОДАТОК Б Титульна сторінка

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
КАФЕДРА ІНФОРМАТИКИ ТА ПРОГРАМНОЇ ІНЖЕНЕРІЇ

КУРСОВА РОБОТА

з дисципліни «Аналіз даних в інформаційних системах»

(назва дисципліни)

на тему: _____

Студента (ки) 2 курсу _____ групи

Спеціальності 121

(прізвище та ініціали)

«ПРИЙНЯВ» з оцінкою

доц. Ліхоузова Т.А. / доц. Олійник Ю.О.

Підпис

Дата

Київ - 2022 рік

ДОДАТОК В Лист завдання

Національний технічний університет України "КПІ"

Кафедра інформатики та програмної інженерії

Дисципліна Аналіз даних в інформаційних системах

Спеціальність 121 "Інженерія програмного забезпечення"

Курс 2 Група ІП-XX

Семестр 4

ЗАВДАННЯ на курсову роботу студента

(прізвище, ім'я, по батькові)

1. Тема роботи _____

2. Строк здачі студентом закінченої роботи 20.05.2022

3. Вихідні дані до роботи методичні вказівки до курсової роботи (+навести джерела даних)

4. Зміст розрахунково-пояснювальної записки (перелік питань, які підлягають розробці)

(навести зміст пояснювальної записки 1-2 рівні розділів)

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

(за необхідності наводимо)

6. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів курсової роботи	Термін виконання етапів роботи	Підписи керівника, студента
1.			
2.			
3.			
4.			
4.			
5.			
6.			
7.			
8.			
9.			

Студент _____
(підпис)

(прізвище, ім'я, по батькові)

Керівник _____
(підпис)

(прізвище, ім'я, по батькові)

"__" _____ 2022 р.

ДОДАТОК Г Приклади бібліографічного опису

1) Василій Великий. Гомілії / Василій Великий ; [пер. з давньогрец. Л. Звонська]. — Львів : Свічадо, 2006. — 307 с. — (Джерела християнського Сходу. Золотий вік патристики IV—V ст. ; № 14).

2) Коренівський Д. Г. Дестабілізуючий ефект параметричного білого шуму в неперервних та дискретних динамічних системах / Коренівський Д. Г. — К. : Ін-т математики, 2006. — 111 с. — (Математика та її застосування) (Праці / Ін-т математики НАН України ; т. 59).

3) Шкляр В. Елементал : [роман] / Василь Шкляр. — Львів : Кальварія, 2005. — 196, [1] с. — (Першотвір).

4) Богомольний Б. Р. Медицина екстремальних ситуацій [Електронний ресурс]] : навч. посіб. для студ. мед. вузів III—IV рівнів акредитації / Б. Р. Богомольний, В. В. Кононенко, П. М. Чуєв. — 80 Min / 700 MB. — Одеса : Одес. мед. ун-т, 2003. — (Бібліотека студента-медика) — 1 електрон. опт. диск (CD-ROM) ; 12 см. — Систем. вимоги: Pentium ; 32 Mb RAM ; Windows 95, 98, 2000, XP ; MS Word 97-2000.— Назва з контейнера.

5) Розподіл населення найбільш численних національностей за статтю та віком, шлюбним станом, мовними ознаками та рівнем освіти [Електронний ресурс] : за даними Всеукр. перепису населення 2001 р. / Держ. ком. статистики України ; ред. О. Г. Осауленко. — К. : CD-вид-во "Інфодиск", 2004. — 1 електрон. опт. диск (CD-ROM) : кольор. ; 12 см. — (Всеукр. перепис населення, 2001). — Систем. вимоги: Pentium-266 ; 32 Mb RAM ; CD-ROM Windows 98/2000/NT/XP. — Назва з титул. екрану.

6) Бібліотека і доступність інформації у сучасному світі: електронні ресурси в науці, культурі та освіті : (підсумки 10-ї Міжнар. конф. „Крим-2003”) [Електронний ресурс] / Л. Й. Костенко, А. О. Чекмарьов, А. Г. Бровкін, І. А. Павлуша // Бібліотечний вісник — 2003. — № 4. — С. 43. — Режим доступу до журн. : <http://www.nbuv.gov.ua/articles/2003/03klinko.htm>.