

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ І СИСТЕМ

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

до самостійної роботи
з дисципліни «Інтелектуальний аналіз даних»
для здобувачів освітнього ступеня «бакалавр»
зі спеціальностей 122 Комп'ютерні науки та
126 Інформаційні системи та технології
усіх форм навчання

Черкаси
2020

УДК 004.8:004.6](07)
М 54

*Затверджено вченою радою ФІТІС,
протокол № 5 від 17.02.2020 р.,
згідно з рішенням кафедри інформаційних
технологій проектування,
протокол № 8 від 10.01.2020 р.*

Упорядник Єгорова О. В., к.т.н., доцент

Рецензент Лавданський А.О., к.т.н., доцент

М 54 Методичні рекомендації до самостійної роботи з дисципліни «Інтелектуальний аналіз даних» для здобувачів освітнього ступеня «бакалавр» зі спеціальностей 122 Комп'ютерні науки та 126 Інформаційні системи та технології усіх форм навчання [Електронний ресурс] / [упоряд. Єгорова О.В.] ; М-во освіти і науки України, Черкас. держ. технол. ун-т. Черкаси: ЧДТУ, 2020. – 13 с. – Назва з титульного екрану.

Методичні рекомендації спрямовані на закріплення та поглиблене вивчення тем дисципліни, набуття навичок самостійного мислення, використання та розробки методів і засобів інтелектуального аналізу даних здобувачами освітнього ступеня «бакалавр» зі спеціальностей 122 «Комп'ютерні науки» та 126 «Інформаційні системи та технології».

УДК 004.8:004.6](07)

Виробничо-практичне
електронне видання
комбінованого використання

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
до самостійної роботи
з дисципліни «Інтелектуальний аналіз даних»
для здобувачів освітнього ступеня «бакалавр»
зі спеціальностей 122 Комп'ютерні науки та
126 Інформаційні системи та технології
усіх форм навчання

Упорядник: **Єгорова Ольга В'ячеславівна**

В авторській редакції.

ЗМІСТ

ВСТУП.....	4
ОРГАНІЗАЦІЯ САМОСТІЙНОЇ РОБОТИ	5
САМОСТІЙНА РОБОТА № 1 – Попередня підготовка даних	6
САМОСТІЙНА РОБОТА № 2 – Класичні методи прогнозування.....	7
САМОСТІЙНА РОБОТА № 3 – Сучасні модифікації методів кластерного аналізу.....	8
САМОСТІЙНА РОБОТА № 4 – Нечітка on-line сегментація та кластеризація...	9
САМОСТІЙНА РОБОТА № 5 – Факторний аналіз.....	11
САМОСТІЙНА РОБОТА № 6 – Інтелектуальний аналіз даних засобами MS SQL Server	12
ЛІТЕРАТУРА.....	13

ВСТУП

Навчальна дисципліна «Інтелектуальний аналіз даних» належить до циклу дисциплін професійної підготовки здобувачів бакалаврського освітнього рівня зі спеціальностей 122 «Комп'ютерні науки» та 126 «Інформаційні системи та технології».

Предметом вивчення навчальної дисципліни є технології інтелектуального аналізу даних.

Метою викладання навчальної дисципліни навчальної дисципліни «Інтелектуальний аналіз даних» є теоретична та практична підготовка здобувачів вищої освіти у напрямку розв'язання задач обробки великих масивів інформації, проектування інформаційного забезпечення інформаційних систем та розробки сценаріїв можливих дій в умовах невизначеності з використанням інтелектуальних методів обчислень.

Основними завданнями вивчення дисципліни «Інтелектуальний аналіз даних» є забезпечити розуміння і засвоєння здобувачами вищої освіти технологій інтелектуального аналізу даних, підготовка до їх вибору, реалізації і використання при вирішенні прикладних задач, ознайомлення зі станом та перспективами розвитку інтелектуальних методів обчислень як одного з напрямів штучного інтелекту.

Навчальна програма дисципліни містить такі змістовні модулі:

- методи інтелектуального аналізу даних;
- застосування інтелектуальних технологій.

Тематичне планування самостійної роботи здобувачів бакалаврського освітнього рівня із дисципліни «Інтелектуальний аналіз даних» охоплює такі теми:

- попередня підготовка даних;
- класичні методи прогнозування;
- сучасні модифікації методів кластерного аналізу;
- нечітка on-line сегментація та кластеризація;
- факторний аналіз;
- інтелектуальний аналіз даних засобами MS SQL Server.

ОРГАНІЗАЦІЯ САМОСТІЙНОЇ РОБОТИ

Самостійна робота здобувачів вищої освіти безпосередньо впливає на глибину, неформальність та стійкість набутих знань і умінь. Вона є основним способом оволодіння навчальним матеріалом у вільний від аудиторних занять час.

Мета і зміст самостійної роботи здобувачів бакалаврського освітнього рівня із дисципліни «Інтелектуальний аналіз даних» визначені навчальною і робочою програмами цієї дисципліни, методичними матеріалами.

Мета самостійної роботи полягає у забезпеченні розуміння і засвоєння здобувачами бакалаврського освітнього рівня навчальної програми дисципліни у повному обсязі шляхом свідомого закріплення, поглиблення й систематизації набутих теоретичних знань, а також набуття практичних навичок використання сучасних інформаційних систем та технологій під час виконання функціональних обов'язків.

Методичні рекомендації до самостійної роботи здобувачів бакалаврського освітнього рівня із дисципліни «Інтелектуальний аналіз даних» містять шість самостійних робіт, які охоплюють основні теми вказаного курсу. Під час виконання самостійної роботи здобувачі вивчають матеріали окремих тем шляхом опрацювання відповідної літератури, виконують ситуативні завдання, здійснюють підготовку до лабораторних робіт та поточного контролю знань.

Контроль виконання самостійної роботи здійснюється у формі співбесіди, де розглядаються результати виконання ситуативних завдань та рівень засвоєння здобувачем теоретичних знань.

Знання здобувачів бакалаврського освітнього рівня із самостійної роботи оцінюються за наступними критеріями:

- високий понятійний рівень;
- глибина, широта, повнота питання, що висвітлюється;
- осмисленість і усвідомленість затверджених положень теми, яка розкривається здобувачем;
- логічність та послідовність викладу під час відповіді;
- використання та аналіз моделей, методів, засобів, які вивчалися;
- самостійність мислення;
- впевненість в правоті своїх суджень;
- вміння виділяти головне;
- вміння встановлювати міжпредметні та внутрішньопредметні зв'язки;
- вміння робити висновки, показувати перспективу розвитку ідеї або проблеми;
- стилістично грамотна мова.

САМОСТІЙНА РОБОТА № 1

Попередня підготовка даних

Мета роботи: Формування знань щодо попередньої підготовки даних набуття практичних навичок програмування і реалізації програмного продукту.

Порядок виконання роботи

1. Опрацювати питання для самостійного вивчення.
2. Виконати ситуативні завдання.
3. Відповісти на питання для самоконтролю.
4. Пройти співбесіду за результатами виконання самостійної роботи.

Перелік питань для самостійного опрацювання

1. Вибір інформативних ознак в моделях регресії.
2. Вибір інформативних ознак в моделях дискримінантного аналізу.
3. Критерії інформативності ознак.
4. Методи відбору інформативних ознак.

Рекомендована література: [1],[2],[4].

Ситуативне завдання

Розробити комп'ютерну програму для розв'язання задачі синтезу інформативних ознак на основі методу незалежних компонентів.

Теми рефератів

1. Нелінійне відображення багатовимірних даних в просторі нижчої розмірності.
2. Застосування нейронних мереж до відбору інформативних ознак.
3. Стохастичний підхід до відбору ознак.
4. Методи синтезу інформативних ознак.
5. Використання програмних засобів для відбору та синтезу ознак.

Питання для самоконтролю

1. В чому полягає сутність задачі відбору інформативних ознак?
2. Дайте характеристику критеріям інформативності ознак.
3. Опишіть процес відбору інформативних ознак.
4. Що є результатом процесу відбору інформативних ознак?
5. Дайте характеристику методам відбору інформативних ознак.
6. Які проблеми супроводжують використання методів відбору інформативних ознак?
7. Вкажіть особливості використання нейронних мереж для відбору інформативних ознак.

САМОСТІЙНА РОБОТА № 2

Класичні методи прогнозування

Мета роботи: Формування знань методів прогнозування та набуття практичних навичок програмування і реалізації програмного продукту.

Порядок виконання роботи

1. Опрацювати питання для самостійного вивчення.
2. Виконати ситуативні завдання.
3. Відповісти на питання для самоконтролю.
4. Пройти співбесіду за результатами виконання самостійної роботи.

Перелік питань для самостійного опрацювання

1. Узагальнений метод найменших квадратів.
2. Методи оцінювання невідомих параметрів у моделях симулятивних рівнянь.
3. Метод непрямих найменших квадратів.
4. Метод двокрокових найменших квадратів.

Рекомендована література: [1],[3],[4].

Ситуативні завдання

1. Розробити комп'ютерну програму для розв'язання задачі прогнозування даних на основі авторегресійних моделей.
2. Розробити комп'ютерну програму для розв'язання задачі прогнозування даних на основі трендових моделей.

Теми рефератів

1. Дистрибутивно-лагові моделі.
2. Рекурсивні моделі.
3. Трендові моделі та методи оцінювання параметрів нелінійних трендових моделей.
4. Авторегресійні методи прогнозування.

Питання для самоконтролю

1. Дайте характеристику узагальненого методу найменших квадратів.
2. Опишіть алгоритм непрямих методу найменших квадратів.
3. Опишіть алгоритм непрямих методу найменших квадратів
4. Які проблеми супроводжують використання авторегресійних методів прогнозування?
5. Вкажіть особливості використання рекурсивних та трендових моделей.

САМОСТІЙНА РОБОТА № 3

Сучасні модифікації методів кластерного аналізу

Мета роботи: Формування знань сучасних модифікацій методів кластерного аналізу та набуття практичних навичок програмування і реалізації програмного продукту.

Порядок виконання роботи

1. Опрацювати питання для самостійного вивчення.
2. Виконати ситуативні завдання.
3. Відповісти на питання для самоконтролю.
4. Пройти співбесіду за результатами виконання самостійної роботи.

Перелік питань для самостійного опрацювання

1. Алгоритм BIRCH.
2. Алгоритм WaveCluster.
3. Алгоритм CLARA.
4. Алгоритми Clarans, CURE, DBScan.

Рекомендована література: [5],[6],[7].

Ситуативні завдання

1. Розробити комп'ютерну програму для розв'язання задачі кластеризації даних методом BIRCH.
2. Виконати аналіз та розробку монетичного алгоритму кластеризації.

Теми рефератів

1. Дослідження стійкості методу одиночного зв'язку при перетворенні початкових даних.
2. Порівняльний аналіз ефективності ієрархічних агломеративних методів на прикладі розв'язання задачі кластеризації.
3. Технологія визначення оптимального числа кластерів.
4. Кластеризація у неметричних просторах.
5. Дослідження ефективності алгоритму SKAT при розв'язанні задачі кластеризації з «нестійкими» таксонами.

Питання для самоконтролю

1. Які особливості визначають застосування алгоритму BIRCH?
2. Опишіть основні кроки алгоритму WaveCluster.
3. Які проблеми супроводжують використання алгоритму CLARA?
4. Вкажіть переваги та недоліки алгоритмів Clarans, CURE, DBScan?
5. Для розв'язання яких задач доцільно використовувати алгоритми Clarans, CURE, DBScan?

САМОСТІЙНА РОБОТА № 4

Нечітка on-line сегментація та кластеризація

Мета роботи: Формування знань з нечіткої on-line сегментації та кластеризації і набуття практичних навичок програмування і реалізації програмного продукту.

Порядок виконання роботи

1. Опрацювати питання для самостійного вивчення.
2. Виконати ситуативні завдання.
3. Відповісти на питання для самоконтролю.
4. Пройти співбесіду за результатами виконання самостійної роботи.

Перелік питань для самостійного опрацювання

1. Пакетні методи нечіткої сегментації-кластеризації.
2. Адаптивні on-line методи нечіткої ймовірнісної та можливісної кластеризації.
3. Адаптивні on-line методи нечіткої робасної кластеризації.
4. Адаптивна нечітка кластеризація та сегментація часових послідовностей.

Рекомендована література: [6],[7].

Ситуативне завдання

Розробити комп'ютерну програму для розв'язання задачі кластеризації та пошуку зображень за структурними ознаками.

Теми рефератів

1. Багатокаскадна декомпозиція простору даних.
2. Інтервальне оцінювання відеорядів.
3. Просторово-часова сегментація відеорядів.
4. Робасні методи прогнозування та сегментації нестационарних часових послідовностей.

Питання для самоконтролю

1. Дайте характеристику імовірнісного методу нечіткої сегментації-кластеризації.
2. Дайте характеристику можливісного методу нечіткої сегментації-кластеризації.
3. Дайте характеристику імовірнісному адаптивному on-line методу нечіткої кластеризації.
4. Дайте характеристику можливісному адаптивному on-line методу нечіткої кластеризації.

5. Дайте характеристику адаптивних on-line методів нечіткої робасної кластеризації.

САМОСТІЙНА РОБОТА № 5

Факторний аналіз

Мета роботи: Формування знань з теорії факторного аналізу та набуття практичних навичок програмування і реалізації програмного продукту.

Порядок виконання роботи

1. Опрацювати питання для самостійного вивчення.
2. Виконати ситуативні завдання.
3. Відповісти на питання для самоконтролю.
4. Пройти співбесіду за результатами виконання самостійної роботи.

Перелік питань для самостійного опрацювання

1. Основні поняття факторного аналізу.
2. Метод головних компонент.
3. Метод головних факторів.
4. Груповий метод.

Рекомендована література: [1],[4],[8].

Ситуативне завдання

Розробити комп'ютерну програму для розв'язання задачі переходу до нового набору факторів методом головних компонент.

Теми рефератів

1. Сучасні тенденції розвитку методу максимуму правдоподібності.
2. Прикладне використання методу контрастних груп.
3. Дослідження ефективності методу мінімальних залишків для різної кількості знайдених розв'язків.
4. Порівняльний аналіз групового методу та методу α -факторного аналізу.

Питання для самоконтролю

1. Які принципи покладені в основу методів факторного аналізу?
2. Для розв'язання яких задач доцільно використовувати метод факторного аналізу?
3. Дайте характеристику методу головних компонент.
4. Які варіації має класичний метод факторного аналізу?
5. Опишіть алгоритм методу головних факторів.
6. В чому полягає головна відмінність базових припущень методу головних факторів від методу головних компонент?
7. Назвіть особливості застосування методу максимум правдоподібності.
8. В яких випадках доцільно використовувати центроїдний метод?

САМОСТІЙНА РОБОТА № 6

Інтелектуальний аналіз даних засобами MS SQL Server

Мета роботи: Набуття практичних навичок інтелектуального аналізу даних засобами MS SQL Server.

Порядок виконання роботи

1. Опрацювати питання для самостійного вивчення.
2. Виконати ситуативні завдання.
3. Відповісти на питання для самоконтролю.
4. Пройти співбесіду за результатами виконання самостійної роботи.

Перелік питань для самостійного опрацювання

1. SQL Server Management Studio.
2. Написання сценаріїв для об'єктів інтелектуального аналізу даних.
3. Завдання інтелектуального аналізу даних служб Integration Services.
4. Етапи Data Mining у MS SQL Server.

Рекомендована література: [5],[9].

Ситуативне завдання

Написати сценарії для дій ALTER, CREATE і DELETE над об'єктами інтелектуального аналізу даних мовою XML для аналітики (XMLA) в середовищі Management Studio.

Теми рефератів

1. Особливості виконання MS SQL Server на віртуальних машинах.
2. Робота з даними в хмарах.
3. Дослідження ефективності показників продуктивності для кластерів великих даних.
4. Адміністрування бази даних SQL Azure.

Питання для самоконтролю

1. Дайте характеристику SQL Server Management Studio.
2. Опишіть процес написання сценаріїв для об'єктів інтелектуального аналізу даних.
3. Назвіть основні завдання інтелектуального аналізу даних служб Integration Services.
4. Дайте характеристику етапів Data Mining у MS SQL Server.
5. Наведіть приклади проектів інтелектуального аналізу даних у MS SQL Server.

ЛІТЕРАТУРА

1. Снитюк В. Є. Прогнозування. Моделі. Методи. Алгоритм: навч. посібник. Київ: Маклаут, 2008. 364 с.
2. Олійник А. О., Субботін С. О., Олійник О. О. Інтелектуальний аналіз даних : навчальний посібник. Запоріжжя : ЗНТУ, 2012. 278 с.
3. Ганчук А. А., Соловйов В. М., Чабаненко Д. М. Методи прогнозування: навч. посібн. Черкаси: Брама-Україна, 2012. 140 с.
4. Бахрушин В. Є. Методи аналізу даних : навч. посібник для студентів. Запоріжжя: КПУ, 2011. 268 с.
5. Гладун А. Я., Рагушина Ю. В. Data Mining: пошук знань в даних. К.: ТОВ «ВД «АДЕФ Україна», 2016. 452 с.
6. Аналіз та обробка потоків даних засобами обчислюваного інтелекту : монографія / Є. В. Бодянський, Д. Д. Пелешко, О. А. Винокурова, С. В. Машталір, Ю. С. Іванов. Львів: Видавництво Львівської політехніки, 2016. 236 с.
7. Мельник Р. А. Алгоритми та методи опрацювання зображень : навч. посібник. Львів: Видавництво Львівської політехніки, 2017. 220 с.
8. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань : навчальний посібник. Львів: «Магнолія 2006», 2015. 276 с.
9. Анализ данных и процессов : учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод и др.; 3-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с.