

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Волинський національний університет імені Лесі Українки
Факультет інформаційних технологій і математики
Кафедра комп'ютерних наук та кібербезпеки

СИЛАБУС
нормативного освітнього компонента
СУЧАСНІ МЕТОДИ ТА МОДЕЛІ
ОБРОБКИ ДАНИХ
підготовки другого (магістерського) рівня вищої освіти
спеціальності 122 Комп'ютерні науки
освітньо-професійної програми
Комп'ютерні науки та інформаційні технології

Силабус нормативного освітнього компонента “Сучасні методи та моделі обробки даних”
підготовки магістрів, галузі знань 12 Інформаційні технології, спеціальності 122 Комп’ютерні
науки, за освітньою програмою Комп’ютерні науки та інформаційні технології

Розробник:

Мамчич Тетяна Іванівна, кандидат фізико-математичних наук, доцент, доцент кафедри
комп’ютерних наук та кібербезпеки

Погоджено

Гарант освітньо-професійної програми:

Булатецький В.В.

**Силабус освітнього компонента затверджено на засіданні кафедри комп’ютерних наук та
кібербезпеки**

протокол № 1 від 25.09.2024 р.

Гришанович Т. О.

Завідувач кафедри:

І ОПИС ОВІТНЬОГО КОМПОНЕНТА

Найменування показників	Галузь знань, спеціальність, освітня програма, освітній ступінь, ОКР	Характеристик освітнього компонента
		Нормативна
Денна форма навчання	Галузь знань 12 Інформаційні технології, спеціальність 122 Комп'ютерні науки, освітньо-професійна програма Комп'ютерні науки та інформаційні технології освітній рівень бакалавр.	Рік підготовки 1
Кількість годин /кредитів150/5		Семестр 1
		Лекції 28 год.
		Лабораторні 44 год.
		Самостійна робота 68 год.
		Консультації 10 год.
ІНДЗ: нема		Форма контролю: екзамен

II Інформація про викладача

ППП: Мамчич Тетяна Іванівна;

Науковий ступінь: кандидат фізико-математичних наук;

Вчене звання: доцент;

Посада: доцент кафедри комп'ютерних наук та кібербезпеки;

Контактна інформація: Mamchych.Tetyana@vnu.edu.ua

Дні занять: <https://ps.vnu.edu.ua/cgi-bin/timetable.cgi?n=700>

III. Опис освітнього компонента

Анотація курсу. Силабус освітнього компонента складено відповідно до освітньо-професійної програми підготовки магістрів “Комп'ютерні науки та інформаційні технології”. Освітній компонент “Сучасні методи та моделі обробки даних” передбачає вивчення здобувачами освіти основних методів роботи з числовими, порядковими та текстовими даними (Data Mining, Text Mining), використання математичних моделей для задач класифікацій, прогнозування та розпізнавання образів, обробки даних різних типів разом з опануванням відповідних комп'ютерних технологій (програма R). Перелік методів базується на сучасних статистичних підходах, залучаючи деякі алгоритми класифікацій, прогнозування та розпізнавання образів.

Предметом вивчення освітнього компонента є методи, моделі та технології аналізу даних.

Мета навчальної освітнього компонента: формування у слухачів знань, вмінь та навичок з аналізу даних, застосування відповідних інформаційних технологій для обробки даних.

Загальні компетентності

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК05. Здатність вчитися й оволодівати сучасними знаннями.

ЗК07. Здатність генерувати нові ідеї (креативність).

Спеціальні (фахові, предметні) компетентності

СК01. Усвідомлення теоретичних засад комп'ютерних наук.

СК02. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.

СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК04. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень.

СК05. Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення

СК07. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

СК11. Здатність ініціювати, планувати та реалізовувати процеси розробки інформаційних та комп'ютерних систем та програмного забезпечення, включно з його розробкою, аналізом, тестуванням, системною інтеграцією, впровадженням і супроводом.

Програмні результати навчання

РН1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.

РН2. Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур.

РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

РН8. Розробляти та застосовувати математичні методи аналізу даних (включно з великими).

РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

РН16. Виконувати дослідження у сфері комп'ютерних наук.

РН19. Аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій.

Структура освітнього компонента

Змістовий модуль 1. Статистичні методи аналізу.

Змістовий модуль 2. Сучасні комп'ютерні технології для аналізу даних.

Назви змістових модулів і тем	Усього	Лек.	Лабор.	Сам. роб.	Конс.	*Форма контролю/ Бали
Змістовий модуль 1. Статистичні методи аналізу						
Тема 1. Описові статистики. Оцінювання параметрів.		2	2	4		Захист лаб. роботи/2 б.
Тема 2. Програмне забезпечення для обробки даних.		2	6	6		Захист лаб. роботи/4 б.
Тема 3. Візуалізація даних. Технології побудови інфографіки.		2	4	6	1	Захист лаб. роботи/4 б.
Тема 4. Теорія оцінювання. Статистичне виведення.		2	2	6	1	Захист лаб. роботи/2 б.
Тема 5. Кореляційний аналіз.		2	4	6	1	Захист лаб.

Регресійний аналіз.						роботи/4 б.
Тема 6. Багатовимірні методи для класифікації даних: кластерний, факторний, дискримінантний аналіз, множинна регресія.		4	4	6	1	Захист лаб. роботи/4 б.
Модульна контрольна робота 1						30 б
Разом за модулем 1	74	14	22	34	4	50
Змістовий модуль 2. Сучасні комп'ютерні технології для аналізу даних						
Тема 7. Аналіз та прогнозування рядів динаміки. Основні алгоритми у моделях прогнозування часових рядів ARMA, ARIMA.		2	4	6	1	Захист лаб. роботи/3 б.
Тема 8. Статистичний аналіз текстів (Text Mining)		2	4	6	1	Захист лаб. роботи/3 б.
Тема 9. Основи машинного навчання, нейронні мережі		2	4	8	1	Захист лаб. роботи/4 б.
Тема 10. Основи стохастичних методів аналізу (симуляції, стохаст. програмування).		4	4	6		Захист лаб. роботи/4 б.
Тема 11. Метричний підхід до моделювання в соціальних науках.		2	2	6		Захист лаб. роботи/2 б.
Тема 12. Алгоритми роботи з порядковими та мережевими даними.		2	4	6	1	Захист лаб. роботи/4 б.
Модульна контрольна робота 2						30 б
Разом за модулем 2	64	14	22	34	4	50 б.
Види підсумкових робіт	Поточне оцінювання (40 балів) Модуль 1 (30 балів) Модуль 2 (30 балів) Екзамен (60 балів)					
Всього годин/ Балів 150/ 100	150	28	44	68	10	100 б.

Завдання для самостійного опрацювання

№ з/п	Тема	Кількість годин
1	Підготовка до лабораторних робіт.	20
2	Опрацювання лекційного матеріалу. Підготовка до контрольних робіт.	20
3	Систематизація здобутих знань перед екзаменом	10
	Разом	50

IV. Політика оцінювання

Політика щодо академічної доброчесності. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно, а результати раніше зданих робіт анулюються і виконуються повторно у порядку визначеному викладачем. При цьому викладач залишає за собою право змінити завдання.

Комунікаційна політика. Здобувачі вищої освіти повинні мати активовану університетську пошту. Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту, можливе інше (додаткове) джерело комунікації, визначене викладачем для більш оперативного зв'язку зі здобувачами освіти.

Політика щодо перескладання. Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання модулів відбувається із дозволу лектора за наявності поважних причин (наприклад, лікарняний).

Політика щодо оскарження оцінювання. Політика щодо оскарження оцінки. Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку згідно «Положення про порядок і процедури вирішення конфліктних ситуацій у Волинському національному університеті імені Лесі Українки»

Політика щодо відвідування занять. Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, академічна мобільність, які необхідно підтверджувати відповідними документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту. За об'єктивних причин навчання може відбуватись в он-лайн формі за погодженням з керівником курсу та деканом факультету.

Визнання результатів навчання, отриманих у формальній, неформальній освіті. Під час вивчення освітнього компонента можливе визнання результатів навчання отриманих у формальній, неформальній та/або інформальній освіті. Порядок визнання результатів навчання для здобувачів вищої освіти, набутих у: формальній освіті (академічна мобільність здобувачів освіти на території України чи поза її межами, для здобувачів освіти, які переводяться, поновлюються з інших ЗВО (вітчизняних чи іноземних); неформальній та/або інформальній освіті здійснюється згідно «ПОЛОЖЕННЯ про визнання результатів навчання, отриманих у формальній, неформальній та/або інформальній освіті у Волинському національному університеті імені Лесі Українки».

При оцінюванні може бути врахований курс "Аналіз даних та статистичне виведення з програмою R" (30 балів), "Візуалізація даних"(10 балів) на платформі відкритих курсів Prometheus.

Підсумковий контроль

Форма контролю – семестровий екзамен. Оцінювання здійснюється за 100-бальною шкалою. Оцінка включає в себе поточний контроль (нараховується за якісне виконання лабораторних робіт) та підсумковий модульний контроль (нараховується за виконання модульних контрольних робіт та модульних тестових робіт). Максимальна кількість балів, яку може отримати здобувач освіти під час поточного оцінювання за семестр – 40 балів. Підсумковий модульний контроль за семестр включає в себе оцінки за всі модульні контрольні роботи, тестові завдання і складає 60 балів.

Якщо за результатами семестру накопичено не менше 75 балів і здобувач освіти погоджується із цим результатом, то оцінка за семестр може виставлятися без складання іспиту. В іншому випадку здобувач освіти складає іспит; максимальна кількість балів, яку

можна отримати на іспиті – 60 балів при цьому бали за підсумковий модульний контроль анулюються.

Екзамен проходить в письмовій формі. Оцінка за семестр у випадку складання іспиту є сумою балів поточного контролю та балів, отриманих під час іспиту.

Питання, які виносяться на екзамен

1. Типи показників для статистичного аналізу. Приклади.
2. Статистичний розподіл даних, способи задання розподілу.
3. Візуалізація статистичного розподілу. Гістограма, кумулята, графік, полігон частот, різні види діаграм.
4. Числові характеристики рівня та розкиду статистичних показників: середнє, мода, медіана, дисперсія, середнє квадратичне відхилення, стандартна похибка, квантилі, квартилі, децилі, децильний коефіцієнт, коефіцієнт варіації, розмах варіації.
5. Точкові та інтервальні оцінки. Зв'язок точності та надійності довірчих інтервалів. Карті контролю якості.
6. Основні поняття статистичного виведення: перевірка статистичних гіпотез, помилки I і II роду, рівень значущості, потужність критерію.
7. Призначення критеріїв Пірсона χ^2 , Стюдента, Фішера, Шапіро, Колмогорова-Смірнова, ANOVA, MANOVA.
8. Кореляційний аналіз. Коефіцієнти зв'язку для двох показників,
9. Парна та множинна регресія. Метод найменших квадратів.
10. Аналіз та прогнозування динамічних рядів, у тому числі із “сезонними змінами”.
11. Методи ARMA, ARIMA аналізу часових рядів.
12. Кластерний аналіз.
13. Факторний аналіз.
14. Дискримінантний аналіз.
15. Поняття про машинне навчання: розпізнавання образів, регресія та класифікація, алгоритми опорних векторів (SVM). Нейронні мережі (NN). Статистичні процедури Bootstrap.
16. Поняття про стохастичні обчислювальні алгоритми: симуляції, метод Монте-Карло, методи випадкового спуску, бджолиного рою для оптимізації, мурашиний алгоритм (Ant Colony Optimization), бджолиний алгоритм (Artificial Bee Colony Algorithm), генетичні алгоритми.
17. Аналіз текстових даних (Text Mining), контент-аналіз заданого корпусу даних. Сучасні можливості програми R для пошуку та аналізу текстових даних.
18. Деякі методи аналізу порядкових даних: коефіцієнт кореляції Спірмена, знаходження колективних рішень (social choice), метрика Кемені-Снелла.
19. Моделі графів для мережевих даних. Збалансованість у моделях знакових графів. Міра статусу Харарі в моделях орієнтованих графів для ієрархічних структур. Моделі ланцюгів Маркова для інформаційних мереж, еволюція ергодичних ланцюгів.
20. Обчислювальні можливості програми R: обчислення в інтерактивному режимі, вектори та матриці, дата фрейми, створення та виконання скриптів, введення та виведення даних, імпорт та експорт.
21. Використання пакетів у програмі R. Статистичні функції для описової статистики, візуалізації даних (plot, hist), перевірки гіпотез, багатовимірних методів, Text Mining (tm), SVM, NN, часові ряди (ts). Можливості пакета “ggplot2”. Статистичні дані на картографічному матеріалі.
22. Програмне забезпечення для аналізу даних (ознайомчий рівень): SPSS, Statgraphics, Systat, Matlab, бібліотека статистичних пакетів Python. Можливості MS Excel (Open Calc).

Екзамен проходить письмово у вигляді складання звіту з відповідями на теоретичні питання та з результатами виконання практичних завдань за допомогою відповідних комп'ютерних програм (копіювання у звіт виконаних команд).

На обчислення, підготовку відповідей та оформлення звіту відводиться 100 хвилин.

Екзаменаційний білет містить 5 теоретичних питань (це можуть також бути завдання на обчислення без комп'ютерних програм, або завдання тестового типу) та 5 практичних завдань на обчислення за допомогою програми R. Для обчислень кожному здобувачу освіти пропонується індивідуальний набір даних.

Приклад екзаменаційного білета:

Всі питання по 6 балів.

Теоретичні питання (короткі письмові відповіді):

1. Характеристики розсіювання даних числового показника.
2. Лінія регресії для прогнозування та класифікації даних.
3. Метод SVM: призначення та алгоритм.
4. Факторний аналіз: призначення та умови застосовності.
5. Обчислити результат колективного вибору для набору індивідуальних профілів (порядкових даних).

Практичні завдання (виконуються з програмою R):

6. Для заданого набору даних обчислити середнє, стандартне відхилення, дисперсію, децильний коефіцієнт, нижній та верхній квартилі.
7. Для заданого набору даних побудувати гістограму частот, додати графік щільності.
8. Для заданого набору текстових даних знайти матрицю частот для вказаних лексем (термінів),
9. Обчислити коефіцієнт кореляції для двох заданих документів відносно вживання заданого набору термінів (дані із завд. 8).
10. Для заданого набору даних виконати факторний аналіз. Дати інтерпретацію отриманих результатів.

Шкала оцінювання (національна та ECTS)

Шкала оцінювання знань здобувачів освіти з освітніх компонентів, де формою контролю є іспит

Оцінка в балах	Лінгвістична оцінка	Оцінка за шкалою ECTS	
		оцінка	пояснення
90–100	Відмінно	A	відмінне виконання
82–89	Дуже добре	B	вище середнього рівня
75–81	Добре	C	загалом хороша робота
67–74	Задовільно	D	непогано
60–66	Достатньо	E	виконання відповідає мінімальним критеріям
1–59	Незадовільно	Fx	Необхідне перескладання

Рекомендована література та інтернет-ресурси

1. Майборода Р. Комп'ютерна статистика. Київ : ВПЦ «Київський університет», 2023. 496 с.
URL: <https://drive.google.com/file/d/1O6rJ5TOJ5v15gVP5vzLsHzBPCj-BtYqL/view>
2. Болюбаш Н.М. Інтелектуальний аналіз даних: навч. Посіб. / Н.М. Болюбаш. – Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2023. – 320 с.
URL: <https://dspace.chmnu.edu.ua/jspui/bitstream/123456789/1461/1/Болюбаш%20Н.%20М.%20Інтелектуальний%20аналіз%20даних.pdf>
3. Інтелектуальний аналіз даних. Частина 1 / М.В. Талах, В.В. Дворжак – Чернівці: Технодрук, 2022. – 367 с. (641.0Kb).
URL: <https://archer.chnu.edu.ua/xmlui/handle/123456789/6751>
4. Майборода Р.Є. "Методичні рекомендації по курсу "Дескриптивна статистика" для студентів III курсу"., 34 р. – 2020. URL: <https://probability.knu.ua/userfiles/mre/descrtasks.pdf>
5. Bast H., Korzen C., Meyer U., Penschuck M. Algorithms for Big Data, Springer, 2023.
URL: <https://library.oapen.org/handle/20.500.12657/61296>
6. Майборода Р.Є., Сугакова О.В. "Побудова статистичних графіків в R за допомогою пакета ggplot2". КНУ (електорне видання), 77 р. - 2023. URL: <https://drive.google.com/file/d/1LOPmkWYlAMzrHftxeW6pUi9DeLDsrZBP/view>
7. TextMining with R. SlideShare. URL: <http://www.slideshare.net/whitish/textmining-with-r> (date of access: 05.09.2024).
8. R: The R Project for Statistical Computing. R: The R Project for Statistical Computing. URL: <https://www.r-project.org> (date of access: 05.09.2024).
9. Package "timeSeries". URL: <https://cran.r-project.org/web/packages/timeSeries/timeSeries.pdf>
10. Create Elegant Data Visualisations Using the Grammar of Graphics. Create Elegant Data Visualisations Using the Grammar of Graphics • ggplot2. URL: <https://ggplot2.tidyverse.org/> (date of access: 05.09.2024).