

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

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

підготовки здобувачів освіти

першого (бакалаврського) рівня

спеціальності 122 Комп'ютерні науки

освітньо-професійної програми

Комп'ютерні науки та інформаційні технології

Силабус нормативного освітнього компонента “Інтелектуальний аналіз даних”
підготовки бакалаврів, 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:



Гришанович Т. О.

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

протокол № 1 від 30 серпня 2023 р.

Завідувач кафедри:



Гришанович Т. О.

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, спеціальність, освітня програма, освітній ступінь, ОКР	Характеристика навчальної дисципліни
		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 2
Кількість годин /кредитів 120/4		Семестр 4
		Лекції 28 год.
		Лабораторні 34 год.
		Самостійна робота 50 год.
ІНДЗ: <u>нема</u>		Консультації 8 год.
	Форма контролю: <u>екзамен</u>	

II Інформація про

викладача ІІІ: 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: <http://94.130.69.82/cgi-bin/timetable.cgi>

III. Опис освітнього компонента

1. **Анотація курсу.** Силабус освітнього компонента складено відповідно до освітньо-професійної програми підготовки бакалаврів Комп'ютерні науки та інформаційні технології. Освітній компонент "Інтелектуальний аналіз даних" передбачає вивчення студентами основних методів роботи з числовими, порядковими та текстовими даними (Data Mining, Text Mining) разом з опануванням відповідних комп'ютерних технологій (мова програмування R). Перелік методів базується на сучасних статистичних підходах, залучаючи деякі алгоритми класифікацій, прогнозування та розпізнавання образів.

Предметом вивчення освітнього компонента є методи й технології аналізу даних.

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

3. Перелік компетентностей випускника

Загальні компетентності

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

- ЗК3.** Знання та розуміння предметної області та розуміння професійної діяльності.
- ЗК4.** Здатність спілкуватися державною мовою як усно, так і письмово.
- ЗК5.** Здатність спілкуватися іноземною мовою.
- ЗК6.** Здатність вчитися й оволодівати сучасними знаннями.
- ЗК7.** Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- ЗК8.** Здатність генерувати нові ідеї (креативність).
- ЗК10.** Здатність бути критичним і самокритичним.
- ЗК11.** Здатність приймати обґрунтовані рішення.
- ЗК12.** Здатність оцінювати та забезпечувати якість виконуваних робіт.
- ЗК13.** Здатність діяти на основі етичних міркувань.

Спеціальні (фахові, предметні) компетентності

СК1. Здатність до математичного формулювання та досліджування неперервних та дискретних математичних моделей, обґрунтування вибору методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук, аналізу та інтерпретування.

СК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

СК6. Здатність до системного мислення, застосування методології системного аналізу для дослідження складних проблем різної природи, методів формалізації та розв'язування системних задач, що мають суперечливі цілі, невизначеності та ризики.

СК7. Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів.

СК8. Здатність проектувати та розробляти програмне забезпечення із застосуванням різних парадигм програмування: узагальненого, об'єктно-орієнтованого, функціонального, логічного, з відповідними моделями, методами й алгоритмами обчислень, структурами даних і механізмами управління.

СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

Програмні результати навчання

ПРН1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

ПРН4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

ПРН12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.

Структура освітнього компонента

Змістовий модуль 1. Класичні методи статистики.

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. Класичні методи статистики						
Тема 1. Описові статистики. Візуалізація даних.	15	4	4	6	1	Звіт по лаб. роботі /4
Тема 2. Теорія оцінювання. Статистичне виведення.	11	2	4	4	1	Звіт по лаб. роботі /4
Тема 3. Кореляційний аналіз. Регресійний аналіз.	13	4	4	4	1	Звіт по лаб. роботі /4
Тема 4. Багатовимірні методи.	15	4	4	6	1	Звіт по лаб. роботі /4
Разом за модулем 1	54	14	16	20	4	16
Змістовий модуль 2. Сучасні комп'ютерні технології для аналізу даних						
Тема 5. Програмне забезпечення для аналізу даних. Програма R.	10	2	4	4		Звіт по лаб. роботі /4
Тема 6. Аналіз та прогнозування динамічних рядів	7	2	2	2	1	Звіт по лаб. роботі /2
Тема 7. Основи машинного навчання, нейронні мережі	11	2	4	4	1	Звіт по лаб. роботі /4
Тема 8. Статистичний аналіз текстів (Text Mining)	12	2	2	8		Звіт по лаб. роботі /2
Тема 9. Основи стохастичних методів аналізу (симуляції, стохаст. програмування).	11	2	2	6	1	Звіт по лаб. роботі /6

Тема 10. Алгоритми роботи з порядковими та мережевими даними.	15	4	4	6	1	Звіт по лаб. роботі / 6
Разом за модулем 2	66	14	18	30	4	24
Модульна контрольна робота 1						30
Модульна контрольна робота 2						30
Разом	120	28	34	50	8	100

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 студентами.

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 студентів на території України чи поза її межами, для студентів, 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» (<https://cutt.ly/yNUt5Y4>).

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 студент 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 студент погоджується із цим результатом, то оцінка за семестр може виставлятися без складання іспиту. В іншому випадку студент 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, кругова діаграма, діаграма “ящик з вусами”.
4. Числові характеристики рівня та розкиду статистичних показників: середнє, мода, медіана, дисперсія, середнє квадратичне відхилення, стандартна похибка, квантілі, квартилі, децилі, децильний коефіцієнт, коефіцієнт варіації, розмах варіації.
5. Точкові та інтервальні оцінки. Зв'язок точності та надійності довірчих інтервалів.
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. Кластерний аналіз.
12. Факторний аналіз.
13. Дискримінантний аналіз.
14. Поняття про машинне навчання: розпізнавання образів, регресія та класифікація, алгоритми опорних векторів (SVM). Нейронні мережі (NN). Статистичні процедури Bootstrap.
15. Поняття про стохастичні обчислювальні алгоритми: симуляції, метод Монте-Карло, метод випадкового спуску для оптимізації, мурашиний алгоритм (Ant Colony Optimization), бджолиний алгоритм (Artificial Bee Colony Algorithm), генетичні алгоритми.
16. Аналіз текстових даних (Text Mining), контент-аналіз заданого корпусу даних. Сучасні можливості програми R для пошуку та аналізу текстових даних.
17. Деякі методи аналізу порядкових даних: знаходження колективних рішень (social choice), метрика Кемені-Снелла.
18. 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.
19. Обчислювальні можливості програми R: обчислення в інтерактивному режимі, вектори та матриці, дата фрейми, створення та виконання скриптів, введення та виведення даних, імпорт та експорт.
20. Використання пакетів у програмі R. Статистичні функції для описової статистики, візуалізації даних, перевірки гіпотез, багатовимірних методів, Text Mining, SVM, NN.

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 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) та 5 практичних завдань на обчислення за допомогою програми R. Для обчислень кожному студенту пропонується індивідуальний набір даних.

Приклад екзаменаційного білета:

Всі питання по 6 балів.

Теоретичні питання (короткі письмові відповіді):

1. Для заданого набору дискретних даних знайти середнє та медіану без використання комп'ютерних програм.
2. Метод найменших квадратів. Лінія регресії.
3. Метод опорних векторів для класифікації багатовимірних даних.
4. Факторний аналіз: призначення та умови застосовності.
5. Обчислити результат колективного вибору для набору індивідуальних профілів (порядкових даних).

Практичні завдання (виконуються з використанням мови програмування R):

6. Для заданого набору даних обчислити середнє та медіану.
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 побудувати лінійну регресію.

Шкала оцінювання (національна та ECTS)

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	Фх	Необхідне перекладання
------	--------------	----	------------------------

Рекомендована література та інтернет-ресурси

1. Майборода Р.Є. *"Комп'ютерна статистика"*. ВПЦ "Київський університет", 2019. 589 с.
<http://probability.univ.kiev.ua/userfiles/mre/cscolor.pdf>
2. Майборода Р. Багатовимірна дескриптивна статистика. Методичні рекомендації по курсу. КНУ ім. Тараса Шевченка, 2020.
<https://probability.knu.ua/userfiles/mre/descrtasks.pdf>
3. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. К.: ІСЗІ КПІ ім. Ігоря Сікорського, 2018. 300 с.
4. TextMining with R : <http://www.slideshare.net/whitish/textmining-with-r>
5. A. Agresti. *Statistical methods for social sciences*. Boston: Pearson. 2018.
6. The R Project for Statistical Computing <https://www.r-project.org/>
7. Сергеев-Горчинський О.О., Іщенко Г.В. Інтелектуальний аналіз даних: комп'ютерний практикум. КПІ ім. Ігоря Сікорського, 2018. 73 с.
https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf