

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

СИЛАБУС
нормативного освітнього компонента
Інтелектуальний аналіз даних

підготовки бакалаврів

спеціальностей 292 Міжнародні економічні відносини, 122 Комп'ютерні науки

освітньо-професійної програми «Міжнародна економіка стартапів і штучний інтелект»

денної (очної) форми навчання

Луцьк 2024

Силабус освітнього компонента «Інтелектуальний аналіз даних» підготовки бакалаврів галузей знань 29 Міжнародні відносини, 12 Інформаційні технології спеціальностей 292 Міжнародні економічні відносини, 122 Комп'ютерні науки, освітньо-професійної програми «Міжнародна економіка стартапів і штучний інтелект».

Розробник: Мамчич Тетяна Іванівна, кандидат фізико-математичних наук, доцент

Погоджено

Гарант освітньо-професійної програми:



О.А. Корнелюк

Силабус освітнього компонента затверджено на засіданні кафедри комп'ютерних наук та кібербезпеки, протокол № 2 від 25.09.2024 р.

Завідувач кафедри:



Т.О. Гришанович

I. ОПИС ОСВІТНЬОГО КОМПОНЕНТА

Найменування показників	Галузь знань, спеціальність, освітня програма, освітній ступінь	Характеристика навчальної дисципліни
Денна форма навчання	галузі знань 29 Міжнародні відносини 12 Інформаційні технології спеціальності 292 Міжнародні економічні відносини 122 Комп'ютерні науки ОПП «Міжнародна економіка стартапів і штучний інтелект» Бакалавр	Нормативна
Кількість годин/кредитів 120/4		Рік навчання: 2-й
		Семестр: 2-й
		Лекції: 30 год.
		Лабораторні роботи: 38 год.
ІНДЗ: немає		Самостійна робота: 44 год.
		Консультації: 8 год.
	Форма контролю: залік	
Мова викладання		Українська

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 , каб. 501, корпус С.

III. ОПИС ОСВІТНЬОГО КОМПОНЕНТА

1. Анотація курсу

Дисципліна «Інтелектуальний аналіз даних» є важливою складовою підготовки бакалаврів спеціальностей 292 «Міжнародні економічні відносини» та 122 «Комп'ютерні науки» освітньо-професійної програми «Міжнародна економіка стартапів і штучний інтелект» для забезпечення аналітичної бази в їхній освіті.

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 (програма R). Перелік методів базується на сучасних статистичних підходах, залучаючи деякі алгоритми класифікацій, прогнозування та розпізнавання образів.

2. Пререквізити: -.

3. Постреквізити: Штучний інтелект, Навчальна практика з освітнього компонента «Штучний інтелект».

4. Мета і завдання освітнього компонента:

формування у слухачів знань, вмінь та навичок з аналізу даних, застосування відповідних інформаційних технологій для обробки даних.

5. Форми та методи навчання: інтерактивна лекція (вербальні та ілюстративно-демонстраційні методи: розповідь, пояснення, дискусія, презентація, демонстрація, запитання-відповідь, мозковий штурм), лабораторні роботи, самостійна робота з літературою та електронними ресурсами, опитування, консультації.

6. Результати навчання (компетентності)

Інтегральна компетентність (ІК). Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми у галузі комп'ютерних наук, або у процесі навчання, що передбачає застосування теорій та методів інформаційних технологій і характеризується комплексністю та невизначеністю умов.

Загальні компетентності (ЗК)

- 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) компетентності (ФК)

ФК1. 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.

ФК2. 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.

ФК11. 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.

ФК16. Здатність реалізовувати високопродуктивні обчислення на основі хмарних сервісів і технологій, паралельних і розподілених обчислень при розробці й експлуатації розподілених систем паралельної обробки інформації.

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

7. 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. Оцінювання параметрів.		2	2	2		10
Тема 2. Візуалізація даних.		2	4	4	1	10
Тема 3. Теорія оцінювання. Статистичне виведення.		2	2	4	1	10
Тема 4. Кореляційний аналіз. Регресійний аналіз.		2	4	4	1	10
Тема 5. Багатовимірні методи.		6	6	8	1	10
Разом за модулем 1	58	14	18	22	4	50
Змістовий модуль 2. Сучасні комп'ютерні технології для аналізу даних						
Тема 6. Основи машинного навчання, нейронні мережі		4	2	6	1	10
Тема 7. Статистичний аналіз		4	6	4	1	10

текстів (Text Mining)						
Тема 8. Основи стохастичних методів аналізу (симуляції, стохаст. програмування).		4	6	6	1	15
Тема 9. Алгоритми роботи з порядковими та мережевими даними.		4	6	6	1	15
Разом за модулем 2	62	16	20	22	4	50
Види підсумкових робіт	Поточне оцінювання (40 балів) Модуль 1 (30 балів) Модуль 2 (30 балів) Екзамен (60 балів)					
Всього годин/ Балів 120/ 100	120	30	38	44	8	100

IV. ПОЛІТИКА ОЦІНЮВАННЯ

Навчання здійснюється з дотриманням норм Кодексу академічної доброчесності ВНУ (<http://surl.li/aagxg>). У випадку відсутності на заняттях з важливих причин, є можливість відпрацювання семінарів шляхом виконання завдань у письмовій формі та усне відпрацювання під час консультацій. Оцінювання знань і умінь студентів здійснюється за модульно-рейтинговою системою. Дисципліна складається з двох змістовних модулів. Підсумкова оцінка за 100-бальною шкалою складається із сумарної кількості балів за: поточне оцінювання з відповідних тем (40 балів), МКР (60 балів). Після завершення лабораторних робіт студенти виконують 2 підсумкові контрольні роботи у програмі R. Якщо студент виконав усі види навчальної роботи, які визначені програмою навчальної дисципліни та отримав не менше 75 балів, отримана сума балів може бути зарахована як підсумкова за згодою студента. В інших випадках, студент складає іспит у формі контрольної роботи в програмі R.

V. ПІДСУМКОВИЙ КОНТРОЛЬ

Форма підсумкового контролю з дисципліни – залік. Якщо студент виконав усі види навчальної роботи, які визначені програмою навчальної

дисципліни та отримав не менше 75 балів, отримана сума балів може бути зарахована як підсумкова за згодою студента. В інших випадках, студент складає іспит у формі тесту в програмі Office-365 (Forms). Тест для складання іспиту містить 60 питань, тривалість виконання – 2,5 години. Положення про поточне та підсумкове оцінювання знань здобувачів ВНУ імені Лесі Українки – https://ed.vnu.edu.ua/wp-content/uploads/2022/08/2022_Polozh_pro_otzin_%D0%A0%D0%B5%D0%B4_%D1%80%D0%B5%D0%B4%D0%9C%D0%95%D0%94.pdf)

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, квантилі, 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 , Ст'юдента, 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 програмою 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.

VI. ШКАЛА ОЦІНЮВАННЯ

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

Сума балів за всі види навчальної діяльності	Лінгвістична оцінка
90 – 100	Відмінно
82 – 89	Дуже добре
75 - 81	Добре
67 -74	Задовільно
60 - 66	Достатньо
1 – 59	Незадовільно

VII. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ІНТЕРНЕТ-РЕСУРСИ

1. Майборода Р.Є. *"Комп'ютерна статистика. Професійний старт"*. ВПЦ "Київський університет", 2023
<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.
<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)
<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
<https://probability.knu.ua/userfiles/mre/descrtasks.pdf>
5. Bast H., Korzen C., Meyer U., Penschuck M. Algorithms for Big Data, Springer, 2023 <https://library.oapen.org/handle/20.500.12657/61296>
6. TextMining with R : <http://www.slideshare.net/whitish/textmining-with-r>
7. The R Project for Statistical Computing <https://www.r-project.org/>