

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

Луцьк – 2023

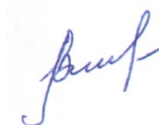
Силабус нормативного освітнього компонента «Сучасні методи та моделі обробки даних» підготовки магістрів, галузі знань 12 – Інформаційні технології, спеціальності 122 – Комп’ютерні науки, за освітньою програмою “Комп’ютерні науки та інформаційні технології”

Розробник:

Собчук Валентин Володимирович, доктор технічних наук, професор, професор 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 30.08.2023 р.

Завідувач кафедри:



Гришанович Т. О.

I. Опис освітнього компонента

Найменування показників	Галузь знань, спеціальність, освітньо-професійна програма, освітній рівень	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: магістр.	Рік підготовки 1
Кількість годин /кредитів 150/5		Семестр 1
		Лекції 32 год.
		Лабораторні 40 год.
ІНДЗ: нема		Самостійна робота 68 год.
	Консультації 10 год.	
		Форма контролю: екзамен

Мова навчання – Українська

II. Інформація про викладача

ППП: Собчук Валентин Володимирович;

Науковий ступінь: доктор технічних наук;

Вчене звання: професор;

Посада: професор кафедри комп'ютерних наук та кібербезпеки;

Контактна інформація: v.v.sobchuk@gmail.com

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 здобувачів другого (магістерського) рівня вищої освіти «Комп'ютерні науки та інформаційні технології». Освітній компонент «Сучасні методи та моделі обробки даних» передбачає вивчення здобувачами 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.

Навчальний курс «Сучасні методи та моделі обробки даних» складає теоретичну основу для опанування взаємно-доповнюючих освітніх компонентів освітньо-професійної програми «Комп'ютерні науки та інформаційні технології» «Технології проектування та розробки вебресурсів», «Архітектура програмного забезпечення», «Реінжиніринг інформаційних систем» та практики з проектування програмного забезпечення.

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

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

Перелік компетентностей здобувача

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. Методи аналізу даних.

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

Назви змістових модулів і тем	Кількість годин					Форма контролю / бали
	Усього	у тому числі				
		Лек.	Лаб.	Сам. роб.	Конс.	
Змістовий модуль 1. Методи аналізу даних						Контрольна робота / 30
Тема 1. Основи інтелектуального аналізу даних.	8	2	2	4		Захист лаб. роботи/2
Тема 2. Алгоритмічне та програмне забезпечення для обробки даних.	14	2	6	6	1	Захист лаб. роботи/6
Тема 3. Візуалізація даних. Технології побудови інфографіки.	13	2	4	6	1	Захист лаб. роботи/4
Тема 4. Математичні методи аналізу інформаційних моделей.	11	2	2	6	1	Захист лаб. роботи/2
Тема 5. Кореляційний аналіз. Регресійний аналіз.	13	4	2	6	1	Захист лаб. роботи/2
Тема 6. Багатовимірні методи для класифікації даних: кластерний, факторний, дискримінантний аналіз, множинна регресія, багатовимірне шкалювання.	15	4	4	6	1	Захист лаб. роботи/4
Разом за змістовим модулем 1	75	16	20	34	5	50

Змістовий модуль 2. Сучасні комп'ютерні технології для аналізу даних						Контрольна робота / 30
Тема 7. Аналіз та прогнозування рядів динаміки. Основні алгоритми у моделях прогнозування часових рядів ARMA, ARIMA.	13	2	4	6	1	Захист лаб. роботи/4
Тема 8. Розпізнавання візуальних об'єктів у системах комп'ютерного зору.	13	2	4	6	1	Захист лаб. роботи/4
Тема 9. Основи машинного навчання, нейронні мережі	13	2	2	8	1	Захист лаб. роботи/2
Тема 10. Основи стохастичних методів аналізу (симуляції, стохаст. програмування).	15	4	4	6	1	Захист лаб. роботи/4
Тема 11. Метричний підхід до моделювання в соціальних науках.	10	2	2	6		Захист лаб. роботи/2
Тема 12. Алгоритми роботи з порядковими та мережевими даними.	15	4	4	6	1	Захист лаб. роботи/4
Разом за змістовим модулем 2	75	16	20	34	5	50
Всього годин/Балів	150	32	40	68	10	150 год. / 100 балів

Завдання для самостійного опрацювання

№ з/п	Тема	Кількість годин
1	Підготовка до лабораторних робіт.	20
2	Опрацювання лекційного матеріалу. Підготовка до контрольних робіт.	16
3	Самостійне опрацювання матеріалу, підготовка доповідей та презентацій.	24
4	Систематизація здобутих знань перед екзаменом	8
	Разом	68

Завдання самостійної роботи, досліджень та проєктної роботи

№ з/п	Тема
1.	Архітектурні принципи паралелізму. Конвеєрні і векторні обчислення.
2.	Багато процесорні обчислювальні системи із загальною і розподіленою пам'яттю.
3.	Багатоядерні архітектури, графічні обчислювачі. Проблеми синхронізації даних. Схеми комутації, їх типові топології
4.	Методи передачі даних. Алгоритми маршрутизації.
5.	Головні характеристики Big Data. Життєвий цикл аналітики даних. Метадані
6.	Порівняльний аналіз сучасних методів обробки великих даних.
7.	Соціальні наслідки використання сучасних методів обробки великих даних.
8.	Етичні аспекти використання сучасних методів обробки великих даних.
9.	Системи управління великими даними. Поточкові дані. SQL.
10.	Data Ingestion. Data Staging. Analysis Layer. Consumption Layer
11.	Методи та моделі прийняття управлінських рішень.
12.	Моделі паралельного програмування. Парадигми паралельного програмування: паралелізм даних і паралелізм завдань.
13.	Порівняльні характеристики багатозадачних, паралельних і розподілених обчислень.

14.	Показники ефективності паралельного програми: прискорення і ефективність.
15.	Методи аналізу паралельних алгоритмів.
16.	Мови програмування використовуються для роботи з фреймворками даних.
17.	Апаратне забезпечення для обробки великих даних.
18.	Характеристики центрів обробки великих даних.
19.	Механізми забезпечення безпеки інформації відповідно до міжнародних стандартів ISO 7498, ISO/IEC 10181.
20.	Огляд сучасних методів машинного навчання
21.	Аналіз застосування сучасних методів машинного навчання в розпізнаванні образів.
22.	Оцінка перспектив розвитку сучасних методів машинного навчання
23.	Соціальні наслідки використання сучасних методів машинного навчання
24.	Застосування штучного інтелекту для аналізу великих даних.
25.	Загрози розвитку штучного інтелекту хибними та/або небезпечними траєкторіями.

IV. Політика оцінювання

Політика викладача щодо студента. Усі учасники освітнього процесу повинні дотримуватись вимог чинного законодавства України, Статуту і Правил внутрішнього розпорядку ВНУ імені Лесі Українки, загальноприйнятих моральних принципів, правил поведінки та корпоративної культури; підтримувати атмосферу доброзичливості, відповідальності, порядності й толерантності. Кожен здобувач повинен бути учасником дистанційного курсу: <https://moodle-cs.vnu.edu.ua/course/view.php?id=155>

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. 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. Перескладання контрольних робіт та та тестувань, відбувається із дозволу лектора і тільки за наявності поважних причин (наприклад, лікарняний).

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.

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 під час поточного оцінювання за семестр – 40 балів. Підсумковий

модульний контроль за семестр включає в себе оцінки за всі модульні контрольні роботи 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 здобувач 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. Карти 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 , 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. Математичні методи аналізу інформаційних моделей.
12. Методи ARMA, ARIMA аналізу часових рядів.
13. Кластерний аналіз.
14. Факторний аналіз.
15. Дискримінантний аналіз.
16. Поняття про машинне навчання: розпізнавання образів, регресія та класифікація, алгоритми опорних векторів (SVM). Нейронні мережі (NN). Статистичні процедури Bootstrap.
17. 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.
18. Аналіз текстових даних (Text Mining), контент-аналіз заданого корпусу даних. Сучасні можливості програми R для пошуку та аналізу текстових даних.
19. Деякі методи аналізу порядкових даних: коефіцієнт кореляції Спірмена, знаходження колективних рішень (social choice), метрика Кемені-Снелла.
20. 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.
21. Обчислювальні можливості програми R: обчислення в інтерактивному режимі, вектори та матриці, дата фрейми, створення та виконання скриптів, введення та виведення даних, імпорт та експорт.
22. 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.

23. 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 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 студенту пропонується індивідуальний набір даних.

Приклад завдань екзаменаційного білета:

Теоретичні питання (короткі письмові відповіді):

1. Характеристики розсіяння 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:

6. Для заданого набору даних обчислити середнє, стандартне відхилення, дисперсію, децильний коефіцієнт, нижній та верхній квантілі.
7. Для заданого набору даних побудувати гістограму частот, додати графік щільності.
8. Для заданого набору текстових даних знайти матрицю частот для вказаних лексем.
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)

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою
90 – 100	A	Відмінно
82 – 89	B	Добре
75 – 81	C	
67 – 74	D	Задовільно
60 – 66	E	
1 – 59	Fx	Незадовільно

V.Рекомендована література та інтернет-ресурси

1. Гладун Ф.Я., Рогушина Ю.В. Data Mining: пошук знань в даних: підручник. Київ: ТОВ «ВД «АДЕФ-Україна», 2016. 452 с
2. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с.
3. Майборода Р. Комп'ютерна статистика. Київ : ВПЦ «Київський університет», 2019. 589 с. URL: <http://probability.univ.kiev.ua/userfiles/mre/cscolor.pdf>
4. Майборода Р. Багатомірна дескриптивна статистика. Методичні рекомендації по курсу. Київ : КНУ ім. Тараса Шевченка, 2020. 35 с. URL: <https://probability.knu.ua/userfiles/mre/descrtasks.pdf>

5. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. К. : ІСЗЗІ КПІ ім. Ігоря Сікорського, 2018. 300 с.
6. Пуятін Є.П., Гороховатський В.О., Матат О.О. Методи та алгоритми комп'ютерного зору: навч. посібник. Харків: Компанія СМІТ, 2006. 236 с
7. TextMining with R. *Share and Discover Knowledge on SlideShare*. URL: <http://www.slideshare.net/whitish/textmining-with-r> (date of access: 1.09.2023).
8. R: The R Project for Statistical Computing. *R: The R Project for Statistical Computing*. URL: <https://www.r-project.org> (date of access: 1.09.2023).
9. Package "timeSeries". *The Comprehensive R Archive Network*. URL: <https://cran.r-project.org/web/packages/timeSeries/timeSeries.pdf> (date of access: 1.09.2023).
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: 1.09.2023).
11. Сергеев-Горчинський О.О., Іщенко Г.В. Інтелектуальний аналіз даних: комп'ютерний практикум. Київ : КПІ ім. Ігоря Сікорського, 2018. 73 с. URL: https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf