

СУЧАСНІ МЕТОДИ ОРГАНІЗАЦІЇ ТА АНАЛІЗУ ДАНИХ

ОПИС КУРСУ

Сучасні методи організації та аналізу даних – обов'язкова освітня компонента загальної підготовки магістрів з комп'ютерних наук, вивчення якої дозволить студенту розуміти сучасний стан та новітні тенденції розвитку технологій аналізу даних та видобування знань, отримати знання перспективних напрямків розвитку ефективних та практичних методів інтелектуального аналізу даних, вміння створювати сховища даних для формування та перевірки гіпотез, отримувати, організовувати та готувати дані різної природи та ступеня структурованості для інтелектуального аналізу і машинного навчання. Здобувач, після оволодіння теоретичними основами реалізації та удосконалення алгоритмів обробки даних і витягу знань, за результатами отримання навичок побудови інформаційних систем для інтелектуального аналізу даних та машинного навчання, зможе працювати як інженер-дослідник та аналітик з організації та аналізу даних, а після відповідної спеціалізації, і як спеціаліст з машинного навчання та штучного інтелекту.

Особливістю курсу є розгляд найбільш актуальних напрямків в області аналізу даних, основних методів інтелектуальної обробки даних, підходів та математичних основ побудови ефективних моделей та технологій видобування знань з даних для вирішення прикладних задач прогнозування та підтримки прийняття рішень. Вивчення методів і алгоритмів аналізу даних та ілюстрація їх роботи на прикладах дозволить студентам ефективно їх опанувати та використовувати знання та навички, отримані під час вивчення цього курсу, при розробці програмного забезпечення для інтелектуального аналізу даних. Викладання дисципліни побудовано на основі новітньої інформації щодо стану і прогресу у технологіях, методах та моделях та з урахуванням кращих українських і світових практик. Отримані знання дозволять здобувачу вищої освіти в майбутньому ефективно приймати участь у плануванні, реалізації та розробці програмного забезпечення у сфері інтелектуального аналізу даних та видобування знань.

ВИМОГИ

- базові знання із вищої математики та математичного аналізу, теорії ймовірності та математичної статистики;
- знання щодо вибору архітектури програмного рішення та навички проектування програмного забезпечення, в тому числі реляційних баз даних, побудови запитів на мові SQL;
- знання основ програмування, в тому числі програмування на Python;
- наявність корпоративного облікового запису @mipolytech.education, Microsoft Teams, Word, Excel;
- наявність особистого логіну та паролю в Moodle (для отримання або поновлення слід звернутися до відповідальної особи на факультеті).

САГАЙДА Павло

pavlo.sahaida@mipolytech.education
доктор технічних наук, доцент, фахівець в сфері інтелектуального аналізу даних, обчислювального інтелекту та data science



mip metinvest
polytechnic

Освітній рівень

Магістр

Кількість
кредитів

5,0

Назва кафедри,
яка пропонує
дисципліну

Цифрових
технологій та
проектно-
аналітичних
рішень

ПРОГРАМНІ РЕЗУЛЬТАТИ НАВЧАННЯ

- 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;
- розробляти та застосовувати математичні методи для аналізу інформаційних моделей;
- розробляти математичні моделі та методи, алгоритмічне та програмне забезпечення аналізу даних (включно з великим);
- аналізувати існуючі цифрові технології, проектувати, розробляти та впроваджувати на підприємствах різних галузей економіки системи цифрового інтелекту, використовуючи сучасні знання бізнес-аналізу, методів інтелектуальної обробки даних, моделей та технологій видобування знань предметної області.

ТЕМАТИКА

Змістовий модуль 1. Процес накопичення даних і витягу моделей на основі їх аналізу, розв'язання задач кластеризації, класифікації і регресії. Методи первісної обробки даних, визначення залежностей параметрів та тенденцій їх зміни. Типи даних, вимірювальні шкали, формати зберігання даних та основи роботи з ними. Методи первісної обробки даних та дослідження структури даних. Методи та моделі використання навчальної інформації. Перевірка статистичних гіпотез. Кореляційний і регресійний аналіз даних. Множинний регресійний аналіз. Лінійна множинна регресійна модель. Перевірка адекватності моделі. Основні поняття та методи аналізу даних. Задачі, процеси та стандарти. Упровадження Data Mining, OLAP і сховищ даних у системи підтримки прийняття рішень. Класифікація напрямків використання методів обчислювального інтелекту. Процес, стандарти та інструментальні засоби Data Mining. Поняття інженерії знань, особливості її організації, програмні засоби. Методи багатомірного розвідувального аналізу. Методи класифікації, регресії та пошуку асоціативних правил. Постановка задачі класифікації, її розв'язання та представлення результатів. Методи оцінювання помилок класифікації. Постановка задачі регресії, її розв'язання та представлення результатів. Методи оцінювання помилок регресії. Пошук асоціативних правил та рекомендаційні системи. Оцінка рекомендаційних систем: метрики точності, методи A/B тестування. Змістовий модуль 2. Інтелектуальний аналіз даних: покращення якості моделей та спеціальні методи здобуття знань з даних. Оцінка точності моделей, навчених на даних, шляхи та засоби покращення їх якості. Методи поліпшення якості моделей машинного навчання. Аугментація. Cross-validation. Боротьба з перенавчанням. Боротьба з дисбалансом даних, втому числі класів. Регуляризація моделі. Feature engineering. Визначення важливості ознак при моделюванні та тестуванні. Методи зменшення розмірності простору ознак. Методи Principal Component Analysis, t-SNE та інші. Сучасні підходи до комплексування моделей, підвищення їх ефективності. Методи прогнозування часових рядів. Базові відомості про нейронні мережі та їх використання у задачах аналізу даних та здобуття знань. Глибокі нейронні мережі. Уведення до deep learning. Організація потоку робочих процесів в Data Mining та спеціальні методи здобуття знань: побудова ML-конвеєрів, використання інструментів AutoML, основи MLOps. Основи м' яких обчислень.

ОРГАНІЗАЦІЯ КУРСУ, ФОРМИ ТА МЕТОДИ НАВЧАННЯ

- Освітній процес будується як комбінація лекцій та самостійного вивчення навчального матеріалу на платформі Moodle – з одного боку, та проблемно орієнтованих практичних занять з відпрацювання аналітично-розрахункових навичок і навичок програмування – з іншого.
- Відвідування лекційних занять є бажаним, однак не обов'язковим; від студентів очікується ознайомлення з матеріалом перед лекцією, що дозволить побудувати лекційне заняття у вигляді



сполучення пояснень викладача та обговорення проблемних питань, які виникли при підготовці до лекції.

– Практичні заняття передбачають аналіз умовно змодельованих ситуацій, синтетичних наборів даних і наборів даних, наданих у вільний доступ, розв'язання задач аналізу предметних областей, організації зберігання та обробки даних в різних форматах, аналізу та візуалізації, витягу з даних моделей, розбір реальних кейсів за матеріалами відкритого доступу; їх відвідування є бажаним.

– Від студента потребується виконати індивідуальні завдання та модульні контрольні роботи у терміни, встановлені у розділі «Розподіл балів за контрольними точками та графік їх виконання».

– З урахуванням поточної ситуації від учасників освітнього процесу очікується виконання вимог безпеки при сигналі «Повітряна тривога», санкції за залишення заняття або неявку на заняття не застосовуються.

– Опціонально доступні індивідуальні та групові консультації. З викладачем можна зв'язатися через електронну пошту, в чаті або в персональній розмові в MS Teams.

ПІДХОДИ ДО ОЦІНЮВАННЯ

Розподіл балів за контрольними точками та графік їх виконання

Види контр. точок	Тижні																	Всього
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Робота на практичних заняттях				10			10					10			10			40
Складання індивідуальних завдань								15									15	30
Модульні контрольні роботи							15								15			30
Всього	50							50							100			

Зміст та вимоги до контрольних точок

Назва контрольної точки	Опис контрольної точки, порядок її проходження та отримання балів
Робота на практичних заняттях	Оцінка за роботу на практичному (семінарському) занятті оголошується наприкінці заняття і може бути оскаржена одразу ж. Мах 10 балів: <ul style="list-style-type: none"> – студент дав пряму і релевантну відповідь на поставлене питання з використанням обґрунтованого посилання на теоретичний матеріал демонструє володіння термінологічним апаратом, математичним та алгоритмічним забезпеченням, здатний адаптувати алгоритм або програмну реалізацію під зміни завдання, в т.ч. у вигляді додаткових запитань / зміг стисло формалізувати вербально сутність розв'язуваної задачі, визначити ключові складові виконання практичної роботи, критерії якості застосованого алгоритму або програмної реалізації (7 бали); – оцінка ініціативності у роботі над проблемою, логічності та структурованості відповіді, здатності комунікувати у команді та під впливом негативних факторів, в т.ч. під тиском викладача та/або групи, вміння вести дискусію та бути критичним та самокритичним (3 бали)
Виконання та захист індивідуального завдання	Підготовлене есе (звіт) у вигляді файлу *.docx, або *.pdf розміщується у відповідному розділі дисципліни в Moodle і перевіряється протягом тижня після завершення терміну подачі. Оскарження оцінки може бути здійснене на останньому практичному занятті модуля. Мах 15 балів: <ul style="list-style-type: none"> – студент підготував есе (звіт) за конкретним завданням з аналізу предметних областей, розробки та удосконалення алгоритмів обробки даних

	<p>в різних форматах, їх аналізу та візуалізації, витягу з даних моделей, розробки відповідних програмних компонентів, в якому: правильно визначив метод та алгоритм розв'язання задачі, обґрунтував своє бачення теоретичними концепціями або моделями, виконав необхідні розрахунки в разі потреби, представив висновок або власне бачення розв'язання задачі, оцінив якість технічного або програмного рішення і окреслив можливі перспективи і обмеженість такого рішення; есе (звіт) структуровано, викладено діловим, науковим або публіцистичним стилем українською (5 балів);</p> <ul style="list-style-type: none"> – есе (звіт) містить комплексну, логічну і оригінальну пропозицію розв'язання задачі; використання штучного інтелекту (ШІ) не забороняється, оскільки пропозиції відомих застосунків ШІ суттєво залежать від обміркованої постановки питання і уточнюючих питань; однак в разі, якщо відповідь, отримана з використанням ШІ, не є комплексною або не відповідає за стилем і викладеними позиціями іншим частинам есе (звіт) або завдання, містить очевидно неправдиву інформацію, то оцінка за цим критерієм знижується (5 балів) – студент під час презентації / захисту есе (звіт) демонструє володіння термінологічним апаратом, математичним та алгоритмічним забезпеченням, відповідає на запитання, здатний швидко адаптувати алгоритм або програмну реалізацію під зміни у індивідуальному завданні (5 бали)
Модульні контрольні роботи	<p>МКР виконуються в Moodle під час останнього практичного заняття в модулі за 1 година 20 хв. В разі неявки або неможливості виконання МКР з поважних причин на таке заняття допускається відкриття виконання МКР за погодженням з викладачем в інший час асинхронно. Кількість спроб не обмежується, однак обмеження по часу виконання МКР залишається. Кожна модульна контрольна робота включає блок задач з матеріалу модуля (max 15 балів). Задачі передбачають обґрунтування порядку розв'язання проблем, виконання розрахунків. При розв'язанні задач оцінюється логіка і обґрунтованість розв'язання, правильність запропонованого алгоритму та програмної реалізації.</p>

Додаткові зауваження:

- студент може оскаржити отримані оцінки в порядку, передбаченому Положенням про організацію освітнього процесу ([Нормативні документи : Polytechnic \(metinvest.university\)](#)) та Положенням про політику та процедури врегулювання конфліктних ситуацій ([Академічні політики : Polytechnic \(metinvest.university\)](#))
- оцінки, отримані за роботу на практичних заняттях не можуть бути відпрацьовані або покращені, окрім процедури оскарження, оцінки за інші види поточного контролю можуть бути покращені за індивідуальною домовленістю з викладачем;
- викладач не має права знижувати оцінку за індивідуальне завдання або модульну контрольну роботу, якщо вони не були складені вчасно, однак в разі, якщо така робота була оцінена пізніше, ніж момент завершення теоретичного навчання у семестрі, то відповідна оцінка не враховується у рейтингу здобувачів освіти.

Форма підсумкового контролю. Порядок визначення підсумкової оцінки

Форма підсумкового контролю	Письмовий екзамен за матеріалом курсу
Умови допуску до підсумкового контролю	Сума оцінок за поточний контроль за семестр становить не менше 35 балів; якщо здобувачі освіти в результаті самооцінки академічного прогресу не впевнені, що набрали 35 балів за поточну успішність, складуть іспит на 85 балів і вище, то вони мають підвищити власні результати поточного контролю до прийнятного рівня
Порядок визначення підсумкової оцінки	<p>Для варіанту екзамену:</p> <ul style="list-style-type: none"> – підсумкова оцінка (ПО) визначається як середнє арифметичне поточної успішності з навчальної дисципліни (О) та оцінки, отриманої під час іспиту (І). В разі, якщо оцінка, отримана на іспиті, менше 60 балів, підсумкова оцінка дорівнює оцінці іспиту: $\begin{cases} \text{ПО} = \frac{0 + \text{I}}{2}, & \text{якщо } \text{I} \geq 60 \\ \text{I}, & \text{якщо } \text{I} < 60 \end{cases}$

Порядок проходження екзамену	Екзамен складається в Moodle у визначений розкладом екзаменаційної сесії період; до складу завдань екзамену (100 балів) входять 2 задачі, які передбачають виконання розрахунків за необхідності та розробку алгоритмічного та програмного забезпечення для розв'язання задач, розробки та удосконалення алгоритмів, розробки програмних компонентів для їх реалізації (по 50 балів кожна). Екзамен оцінює ступінь володіння термінологією та засадами технологій обробки даних в різних форматах, їх аналізу та візуалізації, витягу з даних моделей, розробки відповідних програмних компонентів, інших завдань за проблематикою всього курсу. На складання екзамену надається 3 спроби. Порядок оскарження екзаменаційної оцінки визначений у розділі 10 Положення про організацію освітнього процесу (Нормативні документи : Polytechnic (metinvest.university))
------------------------------	--

Відповідність між прийнятими в університеті шкалами оцінки наведена в таблиці

Бальна шкала	Рівні	Характеристика	Традиційні шкали	
			Іспит	Залік
90-100	A	Студент демонструє видатний рівень досягнення запланованих результатів вивчення навчальної дисципліни, що засвідчують його безумовну готовність до подальшого навчання та/або професійної діяльності за фахом	Відмінно	
82-89	B	Студент виявляє вищий за середній рівень досягнення запланованих результатів вивчення навчальної дисципліни та готовності до подальшого навчання та/або професійної діяльності за фахом, в його знаннях або діях присутні незначні помилки	Добре	Залік
75-81	C	Студент виявляє середній рівень досягнення запланованих результатів вивчення навчальної дисципліни та готовності до подальшого навчання та/або професійної діяльності за фахом, в його знаннях або діях присутні деякі значущі помилки		
67-74	D	Студент виявляє задовільний рівень досягнення запланованих результатів вивчення навчальної дисципліни та готовності до подальшого навчання та/або професійної діяльності за фахом, в його знаннях або діях наявні суттєві помилки	Задовільно	
60-66	E	Наявні мінімально достатні для подальшого навчання та/або професійної діяльності за фахом результати вивчення навчальної дисципліни		
35-59	FX	Низка запланованих результатів навчання не досягнуті. Рівень наявних результатів навчання є недостатнім для подальшого навчання та/або професійної діяльності за фахом	Незадовільно	Незалік
0-34	F	Результати навчання відсутні або критично низькі		

ОСОБЛИВІ ПІДХОДИ ДО ВИЗНАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

– В разі, якщо здобувач освіти засвоїв повністю або частково відповідні програмні результати навчання під час отримання освіти на попередніх кваліфікаційних рівнях або інших дисциплінах, то кредити та оцінка з даної дисципліни може бути перезарахована в порядку, передбаченому Положенням про організацію освітнього процесу ([Нормативні документи : Polytechnic \(metinvest.university\)](#)). Консультацію з даного питання можна отримати у викладача, куратора або гаранта освітньої програми, завідувача кафедри, за якою закріплено цю дисципліну;

– В разі, якщо здобувач освіти обрав цю дисципліну як дисципліну вільного вибору, не зважаючи на той факт, чи вивчалася вона раніше, оцінка та кредити з цієї дисципліни не перезараховуються;

– В разі, якщо здобувач освіти хотів би самостійно вивчити певні курси з проблематики штучного інтелекту та інтелектуального аналізу даних (наприклад, Coursera, Udemy або інших платформ, в т.ч.

платформ відкритих курсів вітчизняних та/або закордонних університетів), то 1) доцільно звернутися до списку рекомендованих вебресурсів або проконсультуватися з викладачем на предмет релевантності самостійно знайденого освітнього ресурсу програмі дисципліни; 2) в разі успішності опанування такого курсу, яке підтверджується сертифікатом або іншим способом, такому здобувачу у порядку, визначеному Положенням про визнання результатів навчання, набутих у неформальній/інформальній освіті [Нормативні документи : Polytechnic \(metinvest.university\)](#), такі результати можуть бути зараховані замість оцінки з певного виду поточного контролю;

– В разі, якщо здобувач освіти реалізував певний вид наукової роботи (тези, стаття, результативна участь у студентській олімпіаді тощо), то у порядку, визначеному Положенням про визнання результатів навчання, набутих у неформальній/інформальній освіті [Нормативні документи : Polytechnic \(metinvest.university\)](#), такі результати можуть бути зараховані замість оцінки з певного виду поточного або навіть підсумкового контролю; консультацію з питань визнання результатів неформальної та інформальної освіти можна отримати в уповноваженої особи від кафедри, яка викладає дисципліну; перелік таких осіб можна знайти за посиланням [Студентам : Polytechnic \(metinvest.university\)](#).

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. McKinney Wes. Python for Data Analysis. O'Reilly Media, Inc., 2022. 582 p.
2. Géron Au. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 3rd Edition. O'Reilly Media, Inc., 2022. 861 p.
3. Leskovec J., Rajaraman A., Ullman J. Mining of Massive Datasets. 3rd Edition. Cambridge : Cambridge University Press, 2020. 603 p.
4. Russel S., Norvig P. Artificial Intelligence: A Modern Approach. Fourth Edition. Prentice Hall, 2020. 1115 p.
5. Гороховатський В. О., Творошенко І. С. Методи інтелектуального аналізу та оброблення даних : навч. посібник. Харків : ХНУРЕ, 2021. 92 с.
6. Zaki M. J., Wagner M. Jr. Data Mining and Machine Learning: Fundamental Concepts and Algorithms. 2nd Edition. Cambridge University Press, 2020. 776 p.

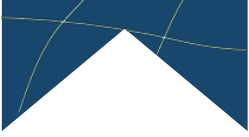
АКАДЕМІЧНІ ПОЛІТИКИ

Як член спільноти Технічного університету «МЕТІНВЕСТ ПОЛІТЕХНІКА» Ви маєте дотримуватися певних стандартів та академічної політики:

– **Академічна недоброчесність** вигляді академічного плагиату; фабрикації; фальсифікації; списування обману; хабарництва; необ'єктивного оцінювання; надання здобувачам освіти під час проходження ними оцінювання результатів навчання допомоги чи створення перешкод, не передбачених умовами та/або процедурами проходження такого оцінювання; впливу у будь-якій формі (прохання, умовляння, вказівка, погроза, примушування тощо) на педагогічного (науково-педагогічного) працівника з метою здійснення ним необ'єктивного оцінювання результатів навчання – прямо заборонено (докладніше про це – у Положенні про академічну доброчесність здобувачів вищої освіти та науково-педагогічних працівників ТОВ ТЕХНІЧНОГО УНІВЕРСИТЕТУ «МЕТІНВЕСТ ПОЛІТЕХНІКА»); і в разі виявлення – **відповідний захід контролю (контрольну точку) буде оцінено в 0 балів за з наступним повідомленням декану факультету та голові комісії з академічної доброчесності Університету.**

– В разі випадку надання здобувачам освіти під час проходження ними оцінювання результатів навчання допомоги чи створення перешкод, не передбачених умовами та/або процедурами проходження такого оцінювання; впливу у будь-якій формі (прохання, умовляння, вказівка, погроза, примушування тощо) на педагогічного (науково-педагогічного) працівника з метою здійснення ним необ'єктивного оцінювання результатів навчання студент може оскаржити процедури оцінювання за процедурами, передбаченими Положенням про організацію освітнього процесу (розділ 10).

– Матеріали в рамках курсу, захищені авторським правом, можуть бути використані лише тільки здобувачами освіти, яким призначено даний курс і для цілей, пов'язаних з цим курсом і не можуть



поширюватися.

- Спілкування з однокурсниками та викладачем має бути професійним та ввічливим.
- Очікується, що Ви перевірятимете всі Ваші письмові повідомлення, включаючи поштові повідомлення та повідомлення у MS Teams на коректність змісту та мови.
- Університет прагне підтримувати середовище, вільне від дискримінації або дискримінаційних домагань, спрямованих на будь-яку людину або групу в межах своєї спільноти - здобувачів освіти, співробітників або відвідувачів.

Докладніше про академічні політики стосовно етичності поведінки, академічної доброчесності та протидію булінгу можна дізнатися за посиланням: [Академічні політики - Polytechnic \(metinvest.university\)](https://metinvest.university)