

ОБРОБКА ВЕЛИКИХ ДАНИХ (BIG DATA)

ОПИС КУРСУ

Обробка великих даних (Big Data) – дисципліна вільного вибору магістерського рівня, вивчення якої дозволить студенту розуміти сучасний стан та новітні тенденції з організації отримання, накопичення та обробка великих даних різної природи з використанням СУБД та мов програмування у складі інформаційних систем, а також спеціалізованих програмних систем, їх ефективного використання для здобуття з великих даних моделей та закономірностей, тобто знань про функціонування предметних областей. Освоєння ефективних практик розробки програмних компонентів для обробки та аналізу великих даних, візуалізації та генерації звітності, дозволить здобувачам вищої освіти у майбутньому працювати як професіонал-аналітик з розробки інформаційних систем, які призначені для опрацювання великих даних.

Особливістю курсу є розгляд сучасних теоретичних здобутків та методик отримання, накопичення та обробки великих даних з використанням спеціалізованих СУБД та мов програмування, які мають у своєму складі відповідні бібліотеки функцій та алгоритмічні засоби, з практичною реалізацією глибокого аналізу великих даних різної природи з видобутком прихованих залежностей та моделей функціонування предметних областей, які ці дані породжують. Студент отримає потрібні компетентності щодо використання сучасних систем управління базами даних масштабу підприємства, наприкладі СУБД MongoDB, Neo4J, MS SQL Server, та засобів розробки програмних комплексів, наприкладі Python, які реалізують багатоаспектну обробку даних для потреб користувачів та забезпечують підтримку прийняття рішень.

Усі напрямки предметної області дисципліни розглянуті в достатньому обсязі для розуміння і подальшого використання на практиці. Отримані компетентності дозволять здобувачу вищої освіти ефективно приймати участь у плануванні, реалізації та розробці інформаційного і програмного забезпечення у сфері комп'ютерних наук та цифрового інтелекту, яке дозволить забезпечити інформаційну підтримку діяльності спеціалістів, в тому числі підтримку прийняття рішень.

Освітній рівень

Магістр

Кількість
кредитів

5,0

Назва кафедри,
яка пропонує
дисципліну

Цифрових
технологій та
проектно-
аналітичних
рішень

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; програмування, в тому числі об'єктно-орієнтованого;
- 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



ПРОГРАМНІ РЕЗУЛЬТАТИ НАВЧАННЯ

- застосовувати 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 – при розробці баз даних з різними моделями організації і зберігання та програмних компонентів для їх обробки;
- розробляти концептуальні моделі, проектувати та супроводжувати бази даних та знань для обробки великих даних, з урахуванням потреб потенційних замовників щодо автоматизації обробки інформації та підтримки прийняття рішень;
- вміти аналізувати існуючі цифрові технології, проектувати, розробляти та впроваджувати на підприємствах різних галузей економіки інформаційні системи, використовуючи сучасні знання бізнес-аналізу, методів інтелектуальної обробки даних, моделей та технологій видобування знань предметної області – при розробці баз даних та програмних компонентів для їх обробки;
- використовувати сучасні системи управління базами даних масштабу підприємства та відповідні системи розробки програмних комплексів для створення компонентів інформаційних систем, які реалізують багатоаспектну обробку даних для потреб користувачів та забезпечують підтримку прийняття рішень;
- проявляти здатність до системного та абстрактного мислення, аналізу та синтезу для дослідження складних проблем різної природи, в тому числі в умовах невизначеності та ризиків, генерувати нові ідеї, бути критичним і самокритичним, приймати обґрунтовані рішення, оцінювати та забезпечувати якість виконуваних робіт в предметній області дисципліни, діяти на основі етичних міркувань

ТЕМАТИКА

Змістовий модуль 1. Отримання, накопичення та обробка великих даних з використанням СУБД та мов програмування. Поняття та термінологія предметної області обробки великих даних. Типи та формати структурованих та погано структурованих даних, методи та засоби їх формалізації. Можливості обробки та аналізу великих даних з використанням мов програмування і їх бібліотек (на прикладі Python). Системи управління реляційними та нереляційними (NoSQL) базами даних, моделі зберігання в таких даних, функціональні можливості, переваги та недоліки їх використання для зберігання та обробки великих даних. Інформаційне та даталогічне моделювання в рамках різних моделей даних. Документо-орієнтовані бази даних та системи управління ними на прикладі СУБД MongoDB. Проектування баз даних у форматі JSON, особливості використання предметно-орієнтованої мови запитів. Графові бази даних, моделі зберігання графів. Системи управління графовими даними на прикладі СУБД Neo4J. Особливості використання предметно-орієнтованої мови запитів Cypher. Застосування графів знань у додатках Retrieval-Augmented Generation (RAG) для підвищення ефективності пошуку. **Змістовий модуль 2.** Глибокий аналіз великих даних різної природи. Витягування, перетворення та завантаження (ETL) даних різної природи та форматів зберігання із різних джерел в файлові структури та бази даних. Web-скрейпінг. Потокові дані, методи та засоби їх обробки, алгоритми їх обробки та використання для інтелектуального аналізу даних. Архітектурні моделі Big Data. Технології віртуалізації. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері. Технології Hadoop Big Data. Розподілена обробка MapReduce. Файлова система HDFS. Місце й роль нейронних мереж та deep learning у розв'язанні задач обробки та аналізу великих даних. Архітектура мереж глибокого навчання та завдання, які вони виконують під час такої обробки.

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 інформаційного та даталогічного моделювання предметних областей, отримання раціональних структур БД з різними моделями зберігання, побудови запитів до них, розробки програмних компонентів для обробки великих даних та витягу з них 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	
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	Підготовлене есе (звіт) у вигляді файлу *.docx, або *.pdf розміщується у відповідному розділі дисципліни в Moodle і перевіряється протягом тижня після

індивідуального завдання	<p>завершення терміну подачі. Оскарження оцінки може бути здійснене на останньому практичному занятті модуля.</p> <p>Мах 15 балів:</p> <ul style="list-style-type: none"> – студент підготував есе (звіт) за конкретним завданням з аналізу даних про предметну область, побудови інформаційної та датованої моделі, отримання раціональної структури бази даних різних моделей зберігання, формування запитів до БД на предметно-орієнтованих мовах, написання програми для глибокого аналізу великих даних, 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 обмеження такого рішення; есе (звіт) структуровано, викладено діловим, науковим або публіцистичним стилем українською (5 балів); – 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 (мах 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	Залік 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	<p>Для варіанту заліку:</p> <ul style="list-style-type: none"> – якщо протягом семестру за результатами поточного контролю здобувач освіти набрав менше 60 балів, то під час екзаменаційної сесії йому надається змога

підсумкової оцінки	отримати/покращити власний результат з усіх видів поточного контролю, крім активності на навчальних заняттях; – в разі, якщо протягом семестру за результатами поточного контролю або в процесі покращення власних результатів здобувач освіти набрав більше 60 балів, йому виставляється фактична сума балів і оцінка «залік», в іншому випадку – «незалік».
--------------------	--

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 \(metinvest.university\)](https://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 \(metinvest.university\)](https://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 \(metinvest.university\)](https://metinvest.university).

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Технології оброблення великих даних : конспект лекцій з дисципліни «Технології оброблення великих даних» / уклад. Л. М. Олещенко. Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с.
2. Ситник Н. В., Зінов'єва І. С. Організація баз даних NoSQL : практикум. Київ: КНЕУ, 2022. 167 с.
3. Сагайда П. І., Зорі А. А., Тарасов О. Ф. Організація комп'ютерних систем для інтелектуальної обробки даних на основі опрацювання формалізованих знань : монографія. Краматорськ : ДДМА, 2020. 191 с.
4. Coronel C., Morris S. Database Systems: Design, Implementation, and Management. 14th Edition. Cengage Learning, 2024. 818 p.
5. Matter U. Big Data Analytics. A Guide to Data Science Practitioners Making the Transition to Big Data. CRC Press, 2023. 328 p.

АКАДЕМІЧНІ ПОЛІТИКИ

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

– **Академічна недоброчесність** вигляді академічного плагіату; 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.**

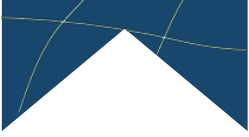
– В разі випадку надання здобувачам освіти під час проходження ними оцінювання результатів навчання допомоги чи створення перешкод, не передбачених умовами та/або процедурами проходження такого оцінювання; впливу у будь-якій формі (прохання, умовляння, вказівка, погроза, примушування тощо) на педагогічного (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)