ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І.І. МЕЧНИКОВА
ФАКУЛЬТЕТ МАТЕМАТИКИ, ФІЗИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА МАТЕМАТИЧНОГО ЗАБЕЗПЕЧЕННЯ КОМП'ЮТЕРНИХ СИСТЕМ

Силабус курсу «Аналіз та візуалізація надвеликих масивів даних (Big Data)»

Обсяг	загальна кількість: кредитів – 4; годин – 120; змістовних модулів – 3
Семестр	весняний
Дні, Час, Місце	за розкладом занять
Викладач(і)	Петрушина Тетяна Іванівна, канд. фіз.-матем. наук, доцент кафедри математичного забезпечення комп'ютерних систем
Контактний телефон	(0674860042)
E-mail	Tatyana.Petrushina@gmail.com
Робоче місце	кафедра математичного забезпечення комп'ютерних систем
Консультації	очні консультації: понеділок з 13.00-14.00 on-line консультації: ZOOM (посилання генерується на початку занять)

КОМУНІКАЦІЯ

Комунікація зі студентами буде здійснюватися електронною поштою, в аудиторії або через ZOOM.

АНОТАЦІЯ КУРСУ

Предметом вивчення курсу є технічні та методологічні засоби організації, зберігання та аналізу надвеликих даних (Big Data).

Пререквізити курсу

Матеріал курсу *ґрунтується* на раніше отриманих студентами знаннях, практичних вміннях та навичках з тем та напрямів щодо алгоритмів, структур даних, реляційних баз даних, мови SQL, дискретной математики та теорії ймовірностей. Відповідні курси викладаються у межах освітньої програми першого (бакалаврського) рівня вищої освіти за спеціальністю 126 «Інформаційні системи та технології».

Постреквізити курсу

Цей курс доповнює дисципліну «Інтелектуальний аналіз даних і методи machine learning» в сфері аналізу та обробки даних і є базою для засвоєння наступних дисциплін освітньо-професійної програми підготовки магістрів за спеціальністю 126 «Інформаційні системи та технології»: «Професійно-дослідницька практика», «Виконання кваліфікаційної роботи магістра».

***Метою* курсу** є формування системних знань щодо організації обчислювальних процесів витягнення та інтерпретації прихованих знань з використанням пакетів прикладних програм різних типів (tensor toolbox, Hadoop, Map reduce та ін.); здатність до розробки інтерфейсного програмного забезпечення для поєднання джерела великих даних (Big Data) з відповідним програмним середовищем; здатність реалізовувати обчислювальний процес на основі хмарних сервісів та технологій, паралельних та розподілених обчислень; набуття знань, практичних навичок та умінь використання сучасних методів, алгоритмів обробки високо вимірних, великооб'ємних, швидко змінюваних різноформатних даних для витягнення прихованих знань, необхідних для підтримки прийняття рішень.

Зміст курсу

Розглядаються:

– *Вступ в аналітику Великих Даних (Big Data). Основні поняття і визначення. Джерела Big Data. Технології формування і обробки Big Data. Застосування Big Data: в економіці, бізнесі, охороні здоров'я, промисловості, приклади використання в наукових областях. Поняття Data Mining, Datafication.*

– *Управління даними, життєвий цикл Big Data. Фази життєвого циклу даних, умовність циклу, принципи формування. Управління даними, 7 головних міркувань у створенні гібридної стратегії інтеграції.*
– *Проблеми уявлення і моделювання знань і їхній зв'язок з проблемою Big Data. Комп`ютерні знання – база даних + база правил. Концептуальні та емпіричні моделі знання. Data Mining - дослідження і виявлення «машиною» в сирих даних прихованих знань.*
– *«Штучний інтелект» як синтез "data mining", "knowledge discovery" і "hidden knowledge". Методи вилучення нових знань з баз фактів. Матричні декомпозиції та їхня роль в аналізі даних.*
– *Багатовимірні масиви, тензорні моделі. Традиційна DataScience - Big Data - загальні риси та відмінності. Тенденція сучасності – від 2D матриць до 3D тензорів. Доцільність і необхідність машинного навчання при опрацюванні результатів обробки Big Data.*
– *База головного понятійного апарат для аналізу Big Data: методи чисельного аналізу на підставі тензорних представлень даних; новітні методи і моделі обробки великих да-них, заснованих на тензорних мережах та тензорних декомпозиціях. Матриці, матричні розкладення.*
– *Інструментальні та програмні засоби роботи з Big Data. Основні принципи роботи з даними; інструменти; приклади розв'язання практичних завдань.*
– *Машинне навчання. Принципи роботи з Big Data, парадигма MapReduce, Hadoop і корпоративні системи, Hadoop і СУРБД.*

ОЧІКУВАНІ РЕЗУЛЬТАТИ

У результаті вивчення курсу студент повинен

знати: основні поняття і термінологію організації, зберігання та аналізу великих даних, технології та сфери застосування інструментів Big Data і методів видобутку даних (Data Mining), розуміти його роль та місце в системах автоматизації прийняття рішень, основні методи, моделі, програмно-технічні засоби, що дозволяють реалізувати обробку багатьох вимірних, швидко-змінюваних та високо-об'ємних масивів даних;

вміти: застосувати набуті знання в професійній діяльності під час розробки, налагодження та експлуатації інформаційних систем з використанням штучного інтелекту та за нестандартних умов проведення дослідження;.

Компетентності, які отримує студент у результаті вивчення курсу:
– *здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.*
– *здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.*
– *здатність застосовувати сучасні моделі та методи нечіткого логічного виводу на підставі форм представлення знань і способів організації пошуку рішень; проектувати та розробляти експертні систем.*
– *здатність проводити інформаційний аналіз та створювати багатовимірні моделі предметних областей.*

Результати навчання: по завершенню курсу студент матиме навички
– *розробляти моделі інформаційних процесів та систем різного класу, використовувати методи моделювання, формалізації, алгоритмізації та реалізації моделей з використанням сучасних комп'ютерних засобів.*
– *розробляти і використовувати інтелектуальні технології для дослідження складних об'єктів, здійснювати аналіз даних для підтримки прийняття рішень.*
– *використовувати сучасні нечіткі моделі, методи та засоби штучного інтелекту в системах прийняття рішень, застосовувати інтелектуальні алгоритми з використанням нечітких моделей для розв'язання задач штучного інтелекту.*
– *створювати оптимізовані пайплайни для підготовки даних до подальшого зберігання та обробки.*

ФОРМИ І МЕТОДИ НАВЧАННЯ

Курс буде викладений у формі лекцій (18 год.) та лабораторних занять (18 год.), організації самостійної роботи студентів (84 год.).

Основна підготовка студентів здійснюється на лекційних та лабораторних заняттях.

Під час викладання курсу використовуються такі **методи навчання**: *словесні* (лекція, пояснення); *наочні* (презентація Power Point); *практичні* (лабораторні роботи); *робота з літературними джерелами* (самостійна робота студентів).

Факультет МФПТ
Кафедра МЗКС