

АНОТАЦІЯ ВИБІРКОВОЇ ДИСЦИПЛІНИ

Назва дисципліни / освітнього компонента	«Інтелектуальний аналіз даних»
Освітня програма	«Інженерія програмного забезпечення»
Компонент освітньої програми	Вибірковий
Загальна кількість кредитів та кількість годин для вивчення дисципліни	3,0 кредити / 90 годин
Вид підсумкового контролю	залік
Мова викладання	українська
Викладач	Сяський Володимир Андрійович, к.т.н., доцент, доцент кафедри інформаційних технологій та моделювання
CV викладача на сайті кафедри	https://kitm.rshu.edu.ua/
E-mail викладача	volodymyr.siasnyi@rshu.edu.ua
Консультації	Згідно з графіком консультацій

МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Навчальна дисципліна «Інтелектуальний аналіз даних» належить до вибірових компонентів циклу професійної підготовки для першого (бакалаврського) рівня вищої освіти за спеціальністю 121 Інженерія програмного забезпечення. Вона вивчається в 6 семестрі після освоєння таких дисциплін: «Математичний аналіз», «Лінійна алгебра і аналітична геометрія», «Дискретний аналіз», «Програмування», «Алгоритми і структури даних», «Математична логіка», «Теорія ймовірностей та математична статистика», «Бази даних та інформаційні системи», «Системний аналіз та методи прийняття рішень», «Аналіз даних».

Інтелектуальний аналіз даних (Data Mining – видобування даних, виявлення знань у базах даних) у сучасному розумінні цього терміну – це процес виявлення в «сирих» даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності. Під терміном «знання» розуміється сукупність відомостей, що утворює цілісний опис, відповідний деякому рівню поінформованості про описуваний об'єкт. Використання отриманих знань означає реальне їх застосування для досягнення конкретних переваг при прийнятті рішень, наприклад, в конкурентній боротьбі за ринок або для створення систем штучного інтелекту.

Суть і мету інтелектуального аналізу даних можна охарактеризувати таким чином: це технологія, призначена для пошуку у великих обсягах даних (*Big Data*) *неочевидних, об'єктивних і корисних на практиці закономірностей*. Неочевидних – тобто таких, що не виявляються стандартними методами обробки інформації або експертним аналізом. Об'єктивних – тобто таких, що будуть повністю відповідати дійсності (на відміну від суб'єктивної експертної думки). Корисних на практиці – тобто таких, яким можна знайти практичне застосування.

Інструменти *Data Mining* можуть і мають знаходити такі закономірності і будувати гіпотези про взаємозв'язки самостійно. Оскільки формулювання гіпотези щодо залежностей є

набагато складнішим завданням, то перевага Data Mining в порівнянні з іншими методами аналізу є очевидною.

Інтелектуальний аналіз даних – потужний засіб теоретичних і практичних досліджень у різних галузях діяльності. Це сфера життєдіяльності, що є одночасно і науковою, оскільки включає в себе наукові дисципліни, і прикладною, оскільки спеціалісту в цій галузі також потрібні знання, вміння і навички в програмуванні та алгоритміці. Крім цього, це в якомусь сенсі культура та мистецтво, оскільки алгоритми Data Mining вимагають постійно шукати нові шляхи вирішення проблем. Тому вивчення такого різновиду інтелектуальних інформаційних технологій є **важливим та актуальним**.

Метою викладання дисципліни «Інтелектуальний аналіз даних» є формування теоретичних та практичних знань про визначальні принципи, методи, технології Data Mining та вміння застосовувати їх при вирішенні прикладних інтелектуальних задач (класифікації, кластеризації, розпізнавання образів, прогнозування, прийняття рішень тощо).

Основними **завданнями** вивчення дисципліни «Інтелектуальний аналіз даних» є:

- засвоєння основних теоретичних положень та стандартів інтелектуального аналізу даних;
- опрацювання передового досвіду використання технологій Data Mining у галузях корпоративного планування, менеджменту, маркетингу, прогнозування, моделювання процесів і систем тощо;
- набуття здатності та навичок програмної реалізації основних алгоритмів інтелектуального аналізу даних та конструювання компонентів систем Data Mining різного призначення і різної проблемної орієнтації із дотриманням принципів, методів та технологій інженерії програмного забезпечення на різних стадіях життєвого циклу інформаційної системи.

ЗМІСТ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. «Основи інтелектуального аналізу даних»

Тема 1. Основні поняття інтелектуального аналізу даних. Цілі та завдання інтелектуального аналізу даних. Поняття даних. Набір даних та їх атрибутів. Вимірювання. Особливості обробки даних. Модель **MapReduce** – модель паралельної обробки великих обсягів даних. Задачі Data Mining. Етапи Data Mining. Методи Data Mining.

Тема 2. Стадії і методи інтелектуального аналізу даних. Класифікація стадій Data Mining. Вільний пошук (Discovery). Прогностичне моделювання (Predictive Modeling). Аналіз винятків (Exception Analysis). Класифікація методів Data Mining. Класифікація методів за принципом роботи з навчальними даними. Класифікація методів за принципами навчання математичних моделей. Класифікація методів за задачами Data Mining. Властивості та особливості застосування методів Data Mining.

Тема 3. Завдання та задачі інтелектуального аналізу даних. Завдання Data Mining. Класифікація задач Data Mining. Рівні аналізу в Data Mining. Дані, інформація й знання.

Змістовий модуль 2. «Задача класифікації даних»

Тема 4. Основні поняття та визначення класифікації. Постановка задачі класифікації. Рівні складності задачі класифікації. Точність класифікації: оцінка рівня помилок.

Тема 5. Методи класифікації даних. Алгоритм **1-rule**. Наївний байєсівський класифікатор. Метод еталонів. Класифікація образів, класи яких частково перетинаються. Метод еталонів, що дробляться. Метод дерев прийняття рішень. Метод лінійної роздільності (лінійної регресії). Метод опорних векторів. Метод найближчого сусіда або системи міркувань «за аналогією».

Тема 6. Застосування штучних нейронних мереж для класифікації. Перцептрони. Лінійні розділяючі відношення. Класифікація за бінарною ознакою. Прошарки перцептронів. Класифікація за **K**-арною ознакою.

Змістовий модуль 3. «Задача кластеризації даних»

- Тема 7. Кластерні технології.** Постановка задачі кластеризації. Точки, простори та відстані. Стратегії кластеризації. Багатовимірні евклідові простори та «прокляття вимірності».
- Тема 8. Ієрархічні алгоритми кластеризації даних.** Базовий алгоритм ієрархічної кластеризації. Алгоритми *Single-link* та *Complete-link*. Альтернативні правила для управління ієрархічною кластеризацією. Ієрархічна кластеризація у неевклідових просторах.
- Тема 9. Неієрархічні алгоритми кластеризації даних.** Алгоритм *k-Means (Hard-c-means)*. Алгоритм Бредлі, Файяда та Рейна (БФР). Метод найближчого сусіда. Алгоритм *CURE*. Неієрархічна кластеризація у неевклідових просторах.
- Тема 10. Застосування штучних нейронних мереж для кластеризації.** Мережі конкурентного неконтрольованого навчання. Кластеризація даних мережами Кохонена.
- Тема 11. Адаптивні методи кластеризації.** Нечіткі алгоритми кластеризації. Генетичні алгоритми кластеризації.
- Тема 12. Асоціативні правила.** Задача пошуку асоціативних правил. Асоціативні правила. Узагальнені асоціативні правила. Чисельні асоціативні правила. Алгоритм *Apriori* та його різновиди.