



Інтелектуальний аналіз даних для задач енергетики

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	Другий (магістерський)
Галузь знань	12 Інформаційні технології
Спеціальність	121 Інженерія програмного забезпечення
Освітня програма	Інженерія програмного забезпечення інтелектуальних кібер-фізичних систем в енергетиці
Статус дисципліни	Нормативна
Форма навчання	Очна (денна)
Рік підготовки, семестр	1 курс, осінній семестр
Обсяг дисципліни	6 кред. /180 год. (лекцій 36год., лабор. 36 год., СРС 108 год.)
Семестровий контроль/ контрольні заходи	Екзамен, МКР
Розклад занять	http://rozklad.kpi.ua/
Мова викладання	Українська/Англійська/Німецька / Французька
Інформація про керівника курсу / викладачів	Лектор/ Лабораторні: д.т.н., ст. науковий співробітник, Верлань Андрій Анатолійович, verlandr@gmail.com , telegram, viber
Розміщення курсу	Кампус

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна «Інтелектуальний аналіз даних» – це важливий напрямок в комп'ютерній науці, що відіграє важливу роль у підготовці фахівців за спеціальністю 121 Інженерія програмного забезпечення. Методи побудування та використання систем штучного інтелекту є невід'ємною частиною програмного забезпечення, які надають машинам інтелектуальних рішень. Крім того методологія, що використовується в цій дисципліні, з успіхом застосовується у звичайних “алгоритмічних” системах, дозволяючи зробити їх більш гнучкими до змін у інформаційному просторі. Вивчаються основи методів інтелектуальної обробки даних. Дисципліна «Інтелектуальний аналіз даних для задач енергетики» (ПО 03) є обов'язковим освітнім компонентом, що належить до циклу професійної підготовки освітньої програми підготовки здобувачів вищої освіти ступеня магістра.

Метою навчальної дисципліни є формування системи знань про типи завдань, що виникають в інтелектуальному аналізі даних (Data Mining), здатності до аналізу великих масивів інформації з метою виявлення нових знань і вмінь, необхідних для прийняття рішень; вивчення основних методів і моделей сучасної обробки даних; формування практичних навичок

роботи із пакетами прикладних програм для розв'язання задач аналізу та інтерпретації даних, розгляд практичних прикладів застосування Data Mining; підготовка студентів до самостійної роботи з вирішення задач засобами Data Mining і розробки інтелектуальних систем.

Предметом навчальної дисципліни є методи і інструменти для побудови систем, що можуть розв'язувати завдання, для яких немає чітких алгоритмів.

Відповідно до освітньої програми підготовки магістрів, вивчення навчальної дисципліни сприяє формуванню наступних компетенцій.

Загальні компетентності:

ЗК 1 Здатність до абстрактного мислення, аналізу та синтезу.

ЗК 3 Здатність проводити дослідження на відповідному рівні.

ЗК 4 Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами інших галузей знань/видів економічної діяльності).

ЗК 5 Здатність генерувати нові ідеї (креативність).

Фахові компетенції:

ФК 5 Здатність розробляти, аналізувати та застосовувати специфікації, стандарти, правила і рекомендації в сфері інженерії програмного забезпечення.

ФК 6 Здатність ефективно керувати фінансовими, людськими, технічними та іншими проектними ресурсами у сфері інженерії програмного забезпечення.

ФК 7 Здатність критично осмислювати проблеми у галузі інформаційних технологій та на межі галузей знань, інтегрувати відповідні знання та розв'язувати складні задачі у широких або мультидисциплінарних контекстах.

ФК 11 Здатність проектувати та розробляти програмні системи з використанням методів інтелектуального аналізу даних.

ФК 13 Здатність реалізовувати застосунки з використанням концепцій інженерії даних та знань.

В той же час, вивчення навчальної дисципліни повинно забезпечити досягнення здобувачами вищої освіти таких **програмних результатів навчання**:

ПРН 1 Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності;

ПРН 2 Оцінювати і вибирати ефективні методи і моделі розроблення, впровадження, супроводу програмного забезпечення та управління відповідними процесами на всіх етапах життєвого циклу.

ПРН 6 Розробляти і оцінювати стратегії проектування програмних засобів; обґрунтовувати, аналізувати і оцінювати варіанти проектних рішень з точки зору якості кінцевого програмного продукту, ресурсних обмежень та інших факторів;

ПРН 7 Аналізувати, оцінювати і застосовувати на системному рівні сучасні програмні та апаратні платформи для розв'язання складних задач інженерії програмного забезпечення;

ПРН 19 Вміти проектувати та розробляти програмні системи з використанням методів інтелектуального аналізу даних;

ПРН 21 Розробляти застосунки з використанням концепцій інженерії даних та знань.

Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Дисципліна «Інтелектуальний аналіз даних» базується на базових знаннях, отриманих при вивченні наступних дисциплін: основи програмування та алгоритмічні мови, а також дискретна математика. У структурно-логічній схемі навчання зазначена дисципліна розміщується тоді, коли студенти вже прослухали курси «Основи програмування»,

«Математичний аналіз», «Лінійна алгебра та аналітична геометрія», «Об'єктно-орієнтований аналіз та конструювання програмних систем», «Комп'ютерна дискретна математика» та набули певного досвіду у програмуванні і можуть виконати складні лабораторні роботи з даної дисципліни.

Отримані в результаті засвоєння дисципліни «Інтелектуальний аналіз даних» теоретичні знання та практичні уміння можуть бути корисні при вивченні курсу «Методологія інженерії програмного забезпечення», а також для проведення наукових досліджень за темою дисертації.

2. Зміст навчальної дисципліни

Зміст навчальної дисципліни повинен забезпечувати виконання мети і всіх завдань робочої програми.

Розділ 1 Загальне уявлення про системи штучного інтелекту в частині Data Mining. Основні задачі та методи Data Mining.

Тема 1. Методи класифікації та прогнозування: дерева рішень, метод опорних векторів, метод найближчого сусіда, метод Naive Bayes.

Тема 2. Метод нейронних мереж: топологія, опис, елементи і архітектура, процес навчання і явище перенавчання. Метод зворотного розповсюдження помилки. Класифікація нейронних мереж, підготовка даних для навчання, вибір структури. Карти Кохонена, що самоорганізуються: задачі, що вирішуються, структура мережі, навчання, алгоритм роботи.

Тема 3. Методи пошуку асоціативних правил: суть задачі, основні характеристики, алгоритм Apriori.

Тема 4. Методи кластерного аналізу. Ієрархічні методи, дендрограма, методи об'єднання або зв'язку кластерів. Ітеративні методи, алгоритм k-середніх, факторний аналіз.

Розділ 2. Нечітка логіка: алгоритми прийняття рішень

Тема 1. Нечіткі алгоритми прийняття рішень на основі *Matlab*.

Тема 2. Побудова нечітких моделей за допомогою субтрактивної кластеризації системи нейро-нечіткого виведення в *Matlab*.

Розділ 3. Генетичні алгоритми, етапи Data Mining–процесу, інтелектуальні системи.

Тема 1. Генетичні алгоритми. Основні поняття. Вирішення задач оптимізації в пакеті *Matlab*.

Тема 2. Етапи Data Mining - процесу. Організаційні та людські фактори в Data Mining. Стандарти Data Mining.. Інструменти Data Mining.

Тема 3. Інтелектуальні системи - базові положення, історія розвитку та сучасні тенденції, філософські аспекти. Різні підходи до побудови ІС. Експертні системи - основні положення.

3. Навчальні матеріали та ресурси

Базові джерела

1. [Paluszek M., Thomas S.](#) Practical MATLAB Deep Learning. 1st Ed.: Apress, 2020. 252p.
2. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2018. 152 с.
3. Олійник А. О., Субботін С. О., Олійник О. О. Інтелектуальний аналіз даних: навчальний посібник, Запоріжжя: ЗНТУ, 2012, 278 с.
4. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: Підручник, К., 2014, 599 с.
5. Sivanandam S. N., Sumathi S., Deera S. N. Introduction to Fuzzy Logic using MATLAB, 2007. 430 p.
6. Сявавко М., Рибицька О. Математичне моделювання за умов невизначеності. — Львів: Українські технології, 2000. — 320 с.

Додаткові джерела

7. Матвійчук А. В. Штучний інтелект в економіці: нейронні мережі, нечітка логіка: Монографія. – К.: КНЕУ, 2011. – 439 с.

8. Матвійчук А. В. Аналіз та прогнозування розвитку фінансово-економічних систем із використанням теорії нечіткої логіки: Монографія / Матвійчук А. В. – К.: Центр навчальної літератури, 2005. – 206 с.

9. Верлань А. А., Гурін А. Л. Системи штучного інтелекту: методичні вказівки до виконання розрахунково – графічної роботи для студентів спеціальності 7.05010105 «Комп’ютерний еколого – економічний моніторинг» денної форми навчання / А. А. Верлань, А. Л. Гурін. – Київ: КПІ, 2012. – 20с.

10. Мороз О. В. Оптимальне управління економічними системами в умовах невизначеності та ризику: Монографія / О. В. Мороз, А. В. Матвійчук. – Вінниця: УНІВЕРСУМ-Вінниця, 2003. – 177с.

11. Потій О. В., Леншин А. В. Основні положення математичного апарату суб’єктивної логіки та його застосування для оцінки рівня зрілості систем забезпечення безпеки інформації // Радіотехніка. Тематичний випуск “Інформаційна безпека”. – Харків: Харківський національний університет радіоелектроніки, 2005. с. 144–160.

12. Zadeh, L. A. Fuzzy logic = computing with words. IEEE Transactions on Fuzzy Systems 4 (2), 1996. p. 103. doi:10.1109/91.493904.

13. Zadeh, L. A. Fuzzy algorithms. Information and Control 12 (2), 1968. p. 94. doi:10.1016/S0019-9958(68)90211-8.

Навчальний контент

4. Методика опанування навчальної дисципліни (освітнього компонента)

Лекційні заняття

№ з/п	Перелік основних питань за темою лекції
	Розділ 1 Загальне уявлення про системи штучного інтелекту в частині Data Mining. Основні задачі та методи Data Mining.
Лекція 1	Основні поняття та визначення інтелектуального аналізу даних (Data Mining). Історія виникнення та причини розвитку. Основні моделі інтелектуальних обчислювань. Суть, мета та сфера застосування технології Data Mining.
Лекція 2	Класифікація задач Data Mining: за призначенням, по способах рішення задачі. Два основні види моделей Data Mining - короткий опис. Методи Data Mining - коротка характеристика. Основні етапи аналізу в процесі виявлення знань. Управління знаннями: класифікація знань, два основні підходи в управлінні знаннями. Засоби Data Mining - категорії, приклади.
Лекція 3	Визначення, різновиди, приклади. Процес класифікації - мета, опис, множини даних, етапи. Методи вирішення задач класифікації. Оцінка точності класифікації, характеристики - критерії оцінки.
Лекція 4	Поняття кластера, алгоритми кластеризації, застосування на практиці. Методи класифікації та прогнозування. Дерева рішень. Методи класифікації і прогнозування - алгоритми дерев рішень. Методи класифікації і прогнозування -

	метод опорних векторів, метод найближчого сусіда. Методи класифікації - Метод Naive Bayes.
Лекція 5	Опис, елементи і архітектура, процес навчання і явище перенавчання нейронної мережі. Моделі нейронних мереж, метод зворотного розповсюдження помилки. Класифікація нейронних мереж, підготовка даних для навчання, вибір структури. Карти Кохонена - задачі, що вирішуються, структура мережі, навчання. Алгоритм роботи мережі Кохонена.
Лекція 6	Суть задачі, основні характеристики, алгоритм Apriori. Методи кластерного аналізу - ієрархічні методи, дендрограма. Методи кластерного аналізу - методи об'єднання або зв'язку кластерів. Методи кластерного аналізу - ітеративні методи, алгоритм k-середніх, факторний аналіз. Порівняльний аналіз ієрархічних і неієрархічних методів кластеризації.
	Розділ 2. Нечітка логіка: алгоритми прийняття рішень
Лекція 7	Основні положення, лінгвістична модель. Нечітка логіка - математичний апарат, функції приналежності. Нечітка логіка - нечіткий логічний вивід. Реалізація нечіткого алгоритму прийняття рішень за допомогою Matlab.
Лекція 8	Синтез нечітких моделей за допомогою субтрактивної кластеризації системи нейро-нечіткого виведення Anfis середовища Matlab.
	Розділ 3. Генетичні алгоритми, етапи Data Mining–процесу, інтелектуальні системи.
Лекція 9	Сфери застосування, основні поняття і компоненти. Опис (схема) роботи генетичного алгоритму. Генетичні алгоритми - основні генетичні оператори. Схема функціонування класичного генетичного алгоритму. Стандартні функції універсального пакета Matlab, призначені для вирішення завдань оптимізації за допомогою генетичних алгоритмів.
Лекція 10	Методи розв'язання задач оптимізації за допомогою ОЛ в пакеті Matlab. Типи генетичних алгоритмів (канонічний, метод переривчастого рівноваги, гібридний алгоритм та ін. Паралельне виконання ГА - паралельний ГА, міграція, глобальна модель, модель дифузії.
Лекція 11	Організаційні та людські фактори в Data Mining. Стандарти Data Mining. Інструменти Data Mining.
Лекція 12	Базові положення, історія розвитку та сучасні тенденції, філософські аспекти. Різні підходи до побудови ІС. Експертні системи - основні положення.

Лабораторні заняття / Модульна контрольна робота / Комп'ютерні практикуми

Основні завдання циклу лабораторних занять / модульної контрольної роботи / комп'ютерних практикумів полягають у набутті студентами практичних навичок з методів інтелектуальної обробки даних та моделювання систем, для яких немає чітких алгоритмів.

№ з/п	Назва теми заняття
-------	--------------------

Лабораторний практикум	
1	Вивчення алгоритмів Data Mining: ознайомлення та навички роботи з GUI інтерфейсом бібліотеки data mining алгоритмів Xelopes, побудова задач пошуку асоціативних правил, кластеризації та класифікації.
2	Виконання аналізу даних методами Data Mining: основні етапи інтелектуального аналізу даних з використанням алгоритмів Data Mining, що реалізовані в бібліотеці Xelopes.
3	Створення програм аналізу даних з використанням алгоритмів Data Mining: реалізація програми, що виконує аналіз даних, які представлені в форматі ARFF за допомогою алгоритму Data Mining та відповідної моделі, заданої варіантом завдання.
4	Робота з нейронними мережами та іншими методами аналітичної обробки даних.
5	Основи аналізу даних: аналіз даних в Microsoft Excel. Застосування апарату кореляційного та регресивності аналізу, описової статистики для вирішення завдань Data Mining
6	Рішення задач класифікації без вчителя (кластеризації) із застосуванням самоорганізованих карт Кохонена.
7	Рішення задачі пошуку асоціативних правил із застосуванням інструментарію Data Mining.
Модульна контрольна робота	
1.1	Рішення задач ієрархічного кластерного аналізу, рішення задач факторного аналізу та ітеративна кластеризація з використанням пакету SPSS.
1.2	Вивчення генетичних алгоритмів - вивчення методів вирішення завдань оптимізації за допомогою GA (Matlab).
Комп'ютерний практикум	
1	Нечіткі алгоритми прийняття рішень на основі Matlab.
2	Побудова нечітких моделей за допомогою субтрактивної кластеризації системи нейро-нечіткого виведення в Matlab.

5. Самостійна робота здобувача вищої освіти

Методика навчання з цієї дисципліни має враховувати не тільки традиційні «аудиторні» засоби навчання, а й засоби дистанційного навчання (ДО). Це стає необхідним, якщо врахувати, що дисципліна викладає матеріал, що є підґрунтям новітніх інформаційних технологій, й тому повинна сама використовувати ці технології в самому процесі навчання. Швидкість розвитку інтелектуальних технологій призводить до того, що моральне старіння наукових розробок, зобов'язує уважно слідкувати за розвитком цієї галузі знань. В цьому питанні основним джерелом інформації може бути мережа Інтернет, з її необмеженими ресурсами. Студенти можуть отримати доступ до сайтів відомих університетів світу, а також до Інтернет-видань. При відсутності виходу в Інтернет корисним може бути використання внутрішньої сітки НТУУ

«КШ». Курс має добре забезпечення електронними посібниками з вивчення інструментарію Data Mining – XELOPS Library, SPSS, пакету Matlab, а також достатньою кількістю навчально-методичних матеріалів на електронних носіях.

№ з/п	Назви тем і питань, що виносяться на самостійне опрацювання та посилання на навчальну літературу
Розділ 1	Огляд застосування методів аналізу інформації (Data Mining) за матеріалами конференцій та періодичних видань.
Розділ 2	Нечітка логіка. Концепція нечіткої множини. Методи визначення функцій приналежності нечітких множин. Методи дефазифікації. Нечіткий кластерний аналіз. Програми для нечіткого моделювання [1,5,7,12,13].
Розділ 3	Різниця між генетичними методами та еволюційним програмуванням. Виділити етапи генетичного алгоритму. Пояснити особливості основних еволюційних алгоритмів [1,4,6].

Політика та контроль

6. Політика навчальної дисципліни (освітнього компонента)

Для успішного проходження курсу та складання контрольних заходів необхідним є вивчення навчального матеріалу за кожною темою. Специфіка курсу передбачає акцент на розумінні підходів і принципів, отримання практичних навичок, а не просто запам'ятовування визначень. Кожен студент повинен ознайомитися і слідувати Положенню про академічну доброчесність, Статуту і розпорядку дня університету. Для успішного засвоєння програмного матеріалу студент зобов'язаний:

- не запізнюватися на заняття;
- не пропускати заняття, а в разі пропуску відновити за допомогою консультування з викладачем та з використанням конспекту на платформі дистанційного навчання «Сікорський», самостійно вивчити матеріал пропущеного заняття та скласти відповідні контрольні заходи в індивідуальному порядку;
- конструктивно підтримувати зворотній зв'язок на всіх заняттях;
- брати активну участь у освітньому процесі;
- своєчасно і старанно виконувати завдання для самостійної роботи;
- бути доброзичливим до однокурсників та викладачів;
- брати участь у контрольних заходах;
- за об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватися індивідуально (в дистанційній online формі за погодженням із деканом факультету);
- будь-яке копіювання або відтворення результатів чужої праці (у тому числі списування), якщо тільки робота не має груповий формат, використання чужих завантажених з Інтернету матеріалів кваліфікується як порушення норм і правил академічної доброчесності та передбачає притягнення винного до відповідальності, у порядку, визначеному чинним законодавством та Положенням про академічну доброчесність університету.

Результатом невиконання та/або недотримання правил може бути оцінка «не зараховано» за курс.

7. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Під час вивчення дисципліни проводиться поточний та підсумковий контроль успішності студентів. Завдання поточного контролю повинні носити практичний характер. Для поточного контролю рекомендується проведення модульної контрольної роботи. Для підсумкового контролю використовується рейтинг студента. Рейтингова оцінка з навчальної дисципліни формується з балів, що отримуються за 7 лабораторних робіт, модульну контрольну роботу та комп'ютерний практикум. Студент може отримати додаткові бали за виконання додаткових завдань або підготовку доповіді на вказану тему. Семестровим контролем є іспит.

1) Лабораторна робота.

Робота складається з практичної та теоретичної частини. Практична частина передбачає написання програмного коду згідно індивідуального завдання. Теоретична частина – відповіді на додаткові питання за темою лабораторної роботи.

2) Модульна контрольна робота.

Модульна контрольна робота складається з 2 задач за темами:

- рішення задач ієрархічного кластерного аналізу, рішення задач факторного аналізу та ітеративна кластеризація з використанням пакету SPSS;
- розв'язання задач оптимізації за допомогою GA (Matlab).

3) Комп'ютерний практикум.

- нечіткі алгоритми прийняття рішень на основі *Matlab*;

- синтез моделей за допомогою субтрактивної кластеризації системи нейро-нечіткого виведення в *Matlab*.

4) Штрафні та заохочувальні бали за:

- виконання додаткових завдань з підвищеною складністю;
- використання в лабораторних роботах самостійно вивчених тем.

Активне навчання - ініціативна самостійна робота в групах.

Створити групи – 2-5 студентів в групі.

Обрати підтему.

Опрацювати / зробити власний огляд за обраною підтемою, використовуючи як рекомендовані, так і будь які інші джерела.

Підготувати коротку доповідь / презентацію до 10 хвилин.

Надати списки груп з темами презентації старостам по мірі готовності. Надалі визначити дати презентацій / дискусій.

Умови допуску до іспиту: зарахування всіх лабораторних робіт, виконання модульної контрольної роботи, комп'ютерних практикумів, а також стартовий рейтинг r_c не менше 40% від $R_{max} = 100$, тобто 40 балів.

На іспиті білет складається з 3 питань: 1 та 2 - теорія за виданими питаннями, 3 - задача за аналогією з лабораторними роботами.

1 частина - підготовка теорії (без комп'ютерів).

2 частина - рішення задачі з комп'ютером.

3 частина - відповіді за білетом (теорія і пояснення рішення задачі).

Рейтинг – $\max 100 = 70$ (\max старт.бал) + 30 (\max залік).

Стартові бали r_c – виходячи з 70 макс. відповідно до успішності та атестації, з урахуванням бонусів та можливим урахуванням мінусів, а також – з урахуванням активності студента та участі у самостійній роботі у групах з підготовкою доповіді/презентації за додатковим матеріалом.

Розмір шкали рейтингових оцінок – 100 балів.

Рейтингова оцінка студента встановлюється за результатами виконання основних (обов'язкових) та додаткових видів робіт.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

Робочу програму навчальної дисципліни (силабус): «Інтелектуальний аналіз даних для задач енергетики»

Складено професором кафедри ІПЗЕ, д.т.н., ст. наук. співробіт. Верланем А.А.

Ухвалено кафедрою ІПЗЕ (протокол № 28 від 15.05.2023р)

Погоджено Методичною комісією НН ІАТЕ¹ (протокол № 9 від 26.05.2023р.)

¹Методичною радою університету – для загальноуніверситетських дисциплін.