



Силабус освітнього компонента Програма навчальної дисципліни



Інтелектуальний аналіз даних

Шифр та назва спеціальності
113 – Прикладна математика

Інститут
ННІ Комп'ютерного моделювання, прикладної фізики та математики

Освітня програма
Комп'ютерне та математичне моделювання

Кафедра
Математичне моделювання та інтелектуальні обчислення в інженерії (161)

Рівень освіти
Бакалавр

Тип дисципліни
Профільна, вільного вибору студента

Семестр
5

Мова викладання
Українська

Викладачі, розробники



Мартиненко Геннадій Юрійович (відповідальний лектор)

gennadii.martynenko@khpi.edu.ua

Доктор технічних наук, професор, професор кафедри математичного моделювання та інтелектуальних обчислень в інженерії НТУ «ХПІ»

Досвід науково-педагогічної роботи – 18 років. Автор понад 180 наукових та навчально-методичних праць.

Лектор та викладач лабораторного практикуму з дисциплін:
«Організація баз даних», «Інтелектуальний аналіз даних», «Програмні комплекси проектування та аналізу», «Програмні засоби моделювання фізичних процесів», «Моделювання об'єктів та процесів в CAD/CAE системах», «Аналіз динамічних процесів в CAD/CAE системах», «Моделювання в CAE системах», «Моделювання нелінійних процесів в CAE системах», «Наближені та чисельні методи розв'язання нелінійних задач», «Педагогічні та інформаційні технології у прикладній математиці».

[Детальніше про викладача на сайті кафедри](#)



Вязовиченко Юлія Андріївна (асистент з лабораторного практикуму)

yuliia.viazovychenko@khpi.edu.ua

Кандидат технічних наук, доцент кафедри математичного моделювання та інтелектуальних обчислень в інженерії НТУ «ХПІ».

Досвід науково-педагогічної роботи – 8 років. Автор понад 25 наукових та навчально-методичних праць.

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна спрямована на формування загальних знань, вмінь та навичок, необхідних для використання технологій інтелектуального аналізу даних (ІАД, англ. Data Mining) та їх реалізації в системах управління базами даних, наприклад, таких як Microsoft SQL Server, PostgreSQL, Greenplum тощо, і пов'язаних програмних продуктах, оволодіння сучасними світовими тенденціями розвитку методів та засобів інтелектуального аналізу даних (вже накопичених при створенні баз даних або в інших джерелах), в поєднанні з інформаційними технологіями для їх застосування в практичній професійній роботі. Розглянуто різні задачі та підходи до створення моделей інтелектуального аналізу даних різними методами та із застосуванням різних алгоритмів, серед яких спрощений алгоритм Байєса, алгоритм дерева прийняття рішень, алгоритм часових рядів, алгоритм кластеризації, алгоритм кластеризації послідовностей, алгоритм взаємозв'язків, алгоритм нейронної мережі, алгоритм лінійної регресії, алгоритм логістичної регресії, та фізичної реалізації цих моделей для аналізу даних з застосуванням мови розширень інтелектуального аналізу даних Data Mining Extensions to SQL (DMX), а також загальну практичну методологію використання цих підходів до моделювання та інструментальних засобів для створення аналізуючих систем в різних областях та галузях.

Мета та цілі дисципліни

Метою викладання дисципліни є: вивчення студентами та формування у них загальних знань за існуючими сучасними підходами до інтелектуального аналізу даних з виконанням всіх типових етапів такого аналізу від постановки задач, підготовки даних, вивчення даних до побудови моделей з їх дослідженням і перевіркою та розгортанням і оновлення цих моделей; набуття вмінь та навичок у застосуванні методів проектування, створення та реалізації моделей інтелектуального аналізу даних з використанням різноманітних спеціалізованих алгоритмів; оволодіння методикою реалізації цих знань за допомогою спеціалізованих систем управління базами даних на практичних прикладах з виконанням групових лабораторних завдань. Цілями викладання дисципліни є: надання студентам поглиблених знань про інтелектуальний аналіз даних; навчання створенню моделей для інтелектуального аналізу даних в різних предметних областях з використанням алгоритмів Байєса, дерева прийняття рішень, часових рядів, кластеризації, кластеризації послідовностей, взаємозв'язків, нейронної мережі, лінійної регресії, логістичної регресії; навчання побудові фізичних моделей, пошуку і аналізу неявних даних у вже накопичених масивах та створенню запитів і прогнозів за допомогою системи управління базами даних, таких як, наприклад, Microsoft SQL Server, PostgreSQL, Greenplum тощо, та мови DMX для загальних та індивідуальних завдань.

Формат занять

Лекції, лабораторні роботи, консультації. Підсумковий контроль – диференційований залік.

Компетентності

згідно освітньої програми:

ЗК10: Навички у використанні інформаційних і комунікаційних технологій;

ФК05: Здатність проектувати бази даних, інформаційні системи та ресурси;

ФК09: Здатність до проведення математичного і комп'ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів;

ФК14: Здатність сформулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв'язання, що забезпечує потрібні точність і надійність результату.

Результати навчання

згідно освітньої програми:

PH06: Володіти основними методами розробки дискретних і неперервних математичних моделей об'єктів та процесів, аналітичного дослідження цих моделей на предмет існування та єдиності їх розв'язку;

PH10: Володіти методиками вибору раціональних методів та алгоритмів розв'язання математичних задач оптимізації, дослідження операцій, оптимального керування і прийняття рішень, аналізу даних;

PH11: Вміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символьних алгоритмів;

PH13: Використовувати в практичній роботі спеціалізовані програмні продукти та програмні системи комп'ютерної математики;

PH19: Збирати та інтерпретувати відповідні дані й аналізувати складності в межах своєї спеціалізації для донесення суджень, які відбивають відповідні соціальні та етичні проблеми.

Обсяг дисципліни

Загальний обсяг дисципліни 4 кредити / 120 год.: лекції – 32 год., лабораторні роботи – 16 год., самостійна робота – 72 год.

Передумови вивчення дисципліни (пререквізити)

Вивчення курсу базується на відомостях (поняття, постановки, підходи та методи програмування, математичного аналізу та логіки, моделювання та обробки даних), що розглядаються в дисциплінах навчального плану:

- ЗП 8. Математичний аналіз;
- ЗП 9. Математичний аналіз;
- СП 4. Лінійна алгебра;
- СП 5. Об'єктно-орієнтоване програмування та проєктування;
- СП 6. Дискретна математика;
- СП 10.. Організація баз даних;
- СП 13. Теорія ймовірності;
- СП 17. Математична статистика;
- СП 11. Математична логіка, теорія алгоритмів та структури даних.

Особливості дисципліни, методи та технології навчання

В рамках курсу «Інтелектуальний аналіз даних» лекції проводяться інтерактивно з використанням мультимедійних технологій. На лабораторних заняттях застосовується практично-орієнтований підхід до навчання, виконуються загальні та індивідуальні завдання, що дозволяє отримати знання та навички у застосування програмних засобів та застосунків для використання різних алгоритмів інтелектуального аналізу даних. При проведенні лабораторного практикуму використовується вільне програмне забезпечення, зокрема РСУБД Microsoft SQL Server (безкоштовний спеціалізований випуск Developer або Express), PostgreSQL, Greenplum тощо.

Програма навчальної дисципліни

Теми лекційних занять

МОДУЛЬ 1. Лекційні заняття (Лк) 2 кредити / 32 год. «ТЕОРЕТИЧНІ ОСНОВИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ (ІАД)»

Тема 1. Базові поняття:

1. Цілі і засоби ІАД; 2. Задачі ІАД; 3. Класифікація задач ІАД.

Тема 2. Інтелектуальний аналіз даних у СУБД Microsoft SQL Server:

1. Реалізація засобів ІАД в СУБД Microsoft SQL Server; 2. Поняття ІАД; 3. Характеристики СУБД MS SQL Server.

Тема 3. Етапи проведення інтелектуального аналізу даних:

1. Схема проведення ІАД; 2. Етапи ІАД.

Тема 4. Короткий огляд алгоритмів ІАД:

1. Спрощений алгоритм Байєса (Naive Bayes); 2. Деревя рішень (Decision Trees); 3. Лінійна регресія (Linear Regression).

Тема 5. Короткий огляд алгоритмів ІАД:

1. Аналіз часових рядів (Time Series); 2. Кластеризація (Clustering).

Тема 6. Короткий огляд алгоритмів ІАД:

1. Алгоритм взаємозв'язків (Association Rules); 2. Кластеризація послідовностей (Sequence Clustering).

Тема 7. Короткий огляд алгоритмів ІАД:

1. Алгоритм нейронних мереж (Neural Network); 2. Алгоритм логістичної регресії (Logistic Regression).

Тема 8. Концепції мови DMX:

1. Базові поняття мови DMX.

Тема 9. Концепції мови DMX:

1. Типи вмісту.

Тема 10. DMX. Створення структури та моделі:

1. Створення структури ІАД.

Тема 11. DMX. Створення структури та моделі:

1. Створення моделі ІАД.

Тема 12. DMX. Обробка, очищення, видалення та відновлення структур і моделей:

1. Обробка даних.

Тема 13. DMX. Обробка, очищення, видалення та відновлення структур і моделей:

1. Обробка даних (продовження); 2. Очищення, видалення та відновлення структур і моделей.

Тема 14. DMX. Робота з даними та побудова прогнозів. Робота з даними:

1. Перегляд варіантів та вмісту; 2. Деталізація структури; 3. Деталізація моделі; 4. Запит значень стовпця; 5. Запит вмісту моделі.

Тема 15. DMX. Робота з даними та побудова прогнозів. Прогнози:

1. Одноелементний запит (прогноз у реальному часі); 2. Запит із використанням функції OPENQUERY; 3. Використання функції Predict; 4. Прогнозування найімовірнішого значення; 5. Створення копії моделі – оператор SELECT INTO.

Тема 16. DMX. Робота з даними та побудова прогнозів. Параметри алгоритмів ІАД. Резюме. Модульний контроль №1.

Теми практичних занять

Немає.

Теми лабораторних робіт

МОДУЛЬ 2. Лабораторні заняття (ЛБ) 1 кредити / 16 год. «ПОБУДОВА МОДЕЛЕЙ ІАД В СУБД» (на одному занятті за 2 год. розглядається по 2 теми)

Тема 1. Надбудови інтелектуального аналізу даних:

1. Встановлення надбудов інтелектуального аналізу; 2. Засіб налаштування сервера; 3. Створення чи вибір бази даних аналітичних служб; 4. Налаштовані з'єднання; 5. Створення нового підключення; 4. Форматування підготовлених даних.

Тема 2. Використання інструментів «Аналіз ключових факторів впливу» та «Виділення категорій»:

1. Вибір залежного параметра для аналізу; 2. Вибір параметрів, від яких аналізується; 3. Звіт, порівняльний звіт; 4. Вибір параметрів аналізу для створення категорій; 5. Налаштування параметрів; 6. Діаграма профілів категорій.

Тема 3. Використання інструментів «Заповнення за прикладом» та «Прогноз»:

1. Підготовка даних для використання інструменту «Заповнення за прикладом»; 2. Алгоритм Logistic Regression; 3. Опис класів задач, для яких придатний алгоритм; 4. Налаштування інструменту «Заповнення за прикладом»; 5. Звіт про виявлені шаблони; 6. Використання алгоритму часових рядів для прогнозу; 7. Опис алгоритму часових рядів; 8. Типи шаблонів; 9. Налаштування параметрів інструменту «Прогноз»; 10. Формування звіту та його інтерпретація.

Тема 4. Використання інструментів «Виявлення винятків» та «Аналіз сценаріїв»:

1. Використання алгоритму кластеризації для виявлення винятків; 2. Вибір параметрів аналізу; 3. Сортування винятків; 4. виправлення помилок; 5. Використання алгоритму Logistic Regression для аналізу сценаріїв; 6. Аналіз сценарію пошуку рішень; 7. Аналіз можливих варіантів; 8. Інтерпретація результатів.

Тема 5. Використання інструментів «Розрахунок прогнозу» та «Аналіз кошика покупок»:

1. Налаштування параметрів інструменту «Розрахунок прогнозу»; 2. Чотири категорії прогнозу; 3. Формування звіту та його інтерпретація; 4. Використання алгоритму Association Rules для аналізу кошика покупок; 5. Вибір параметрів інструменту; 6. Два види звітів та їх інтерпретація.

Тема 6. Використання інструментів Data Mining Client для підготовки даних:

1. Попередній перегляд і представлення даних; 2. Інструмент «Очищення даних» та його налаштування; 3. Перепозначення даних; 4. Інструмент «Зразки даних», особливості використання.

Тема 7. Використання інструментів Data Mining Client для створення моделі інтелектуального аналізу даних:

1. Опис моделей і алгоритмів, які вони використовують; 2. Використання інструменту «Класифікація», дерево рішень; 3. Налаштування параметрів, розподіл даних; 4. Деталізація даних; 5. Діаграма залежностей; 6. Менеджер моделей вкладки Data Mining.

Тема 8. Аналіз точності прогнозу та використання моделі інтелектуального аналізу:

1. Інструмент «Діаграма точності»; 2. Його опис і налаштування; 3. Способи представлення результатів; 4. Аналіз і інтерпретація результатів; 5. Аналіз точності за допомогою інструменту «Матриця класифікації»; 6. Створення моделі для класифікації, що базується на алгоритмі Нейронних мереж; 7. Інструмент «Запит» до моделі Data Mining; 8. Його налаштування і особливості роботи; 9. Конструктор запитів.

Тема 9. Побудова моделі кластеризації, трасування та перехресна перевірка:

1. Інструмент «Кластер»; 2. Налаштування параметрів моделі кластеризації; 3. Вікно перегляду результатів, варіанти представлення (діаграма, профіль кластера, характеристики кластеру, попарне порівняння); 4. Інструмент «Трасування»; 5. Інструмент «Перехресна перевірка»; 6. Вибір моделі, налаштування параметрів, аналіз звіту.

Тема 10. Початок роботи у BIDevStudio:

1. Два режими роботи: негайний та офлайн (опис режимів, застосування, налаштування); 2. Створення нового джерела даних; 3. Створення нового підключення.

Тема 11. Створення представлення джерела даних:

1. Вибір джерела даних; 2. Вибір об'єктів, що включаються до подання; 3. Діаграма зв'язку таблиць; 4. Дослідження даних; 5. Створення іменованих обчислень та запитів.

Тема 12. Створення структури та моделі інтелектуального аналізу. Задача кластеризації:

1. Створення нової структури, параметри налаштування; 2. Вибір алгоритму; 3. Вибір подання джерела даних, таблиці варіантів, вхідних та вихідних атрибутів; 4. Уточнення типів даних; 5. Іменування структури і вибір моделі ІАД; 6. Налаштування під час обробки структури; 7. Усунення помилок при обробці структури; 8. Перегляд побудованої структури.

Тема 13. Задача класифікації. Створення структури та моделей інтелектуального аналізу.

Порівняння точності моделей:

1. Створення структури інтелектуального аналізу та моделі, що використовує спрощений алгоритм Байєса; 2. Дискретизація континуальних атрибутів; 3. Створення додаткових моделей (дерево рішень, нейронна мережа) всередині структури; 4. Порівняльний аналіз моделей за допомогою діаграми точності.

Тема 14. Перегляд моделей інтелектуального аналізу (дерева рішень, спрощений алгоритм Байєса, нейронні мережі). Написання «одноеlementних» прогнозуючих запитів:

1. Використання метаданих; 2. Робота з діаграмою типу «мережа залежностей»; 3. Порівняння наборів значень вхідних атрибутів, що відповідають вибраним значенням вихідного атрибуту; 4. Перегляд моделей; 5. Побудова прогнозів. Вибір моделі; 6. Вибір типу запиту, одноelementний запит; 7. Визначення параметрів, атрибутів і мети запиту.

Тема 15. Робота з моделями інтелектуального аналізу даних із SQL Server Management Studio:

1. Підключення до аналітичних служб SQL Server; 2. Створення запиту DMX; 3. Генерація скрипту на основі існуючого об'єкта (структури чи моделі); 4. Перегляд моделі з Management Studio.

Тема 16. Використання алгоритму Microsoft Time Series для прогнозування значень часових рядів:

1. Алгоритм Time Series; 2. Формування часового ряду за допомогою представлення dbo.vTimeSeries; 3. Ключові атрибути (вхідні і вихідні); 4. Вибір типів даних і вмісту; 5. Зазначення періодичності; 6. Прогнозування і аналіз результатів.

Модульний контроль №2.

Самостійна робота

1. Забезпечення аудиторних занять (опрацювання лекційного матеріалу та оформлення звітів за результатами лабораторних робіт) – 16 год.
2. Самостійне вивчення тем та питань, які не викладаються на лекційних заняттях, за наданими методичними вказівками – 24 год.
3. Забезпечення індивідуальних завдань (виконання за варіантами та оформлення) – 24 год.
4. Забезпечення семестрового контролю (підготовка до модульного контролю) – 8 год.

Література та навчальні матеріали

Основна література

1. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посіб. Харків: ХНУРЕ, 2021. 92 с. (<https://openarchive.nure.ua/server/api/core/bitstreams/...>)
2. Сергеев-Горчинський О.О., Іщенко Г.В. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч. посіб. КПІ ім. Ігоря Сікорського. Київ : КПІ ім. Ігоря Сікорського, 2018. 73 с. (https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf)
3. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2018. 152 с.
4. Олійник А.О., Субботін С.О., Олійник О.О. Інтелектуальний аналіз даних: навч. посіб., Запоріжжя : ЗНТУ, 2012, 278 с.

Додаткова література

1. Майборода Р.Є. "Комп'ютерна статистика". ВПЦ "Київський університет", 2019. 589 с. (<http://probability.univ.kiev.ua/userfiles/mre/cscolor.pdf>)
2. Майборода Р.Є., Сугакова О.В. Аналіз даних за допомогою пакета R. навч. посіб. Київ, КНУ ім. Тараса Шевченка, 2015. 65 с.
3. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навч. посіб. К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. 300 с.
4. Бахрушин В.Є. Методи аналізу даних: навч. посіб. Запоріжжя: КПУ, 2011. 268 с. (<http://kist.ntu.edu.ua/textPhD/metDataManing.pdf>)
5. Мартиненко Г.Ю. Концептуальне та логічне проектування реляційних баз даних [Електронний ресурс] : навч.-метод. посібник. Харків: Нац. техн. ун-т «Харків. політехн. ін-т», 2023. 91 с. (<https://repository.kpi.kharkov.ua/handle/KhPI-Press/70293>)

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Змістовий модуль 1 (Лк) – максимум 50 балів: комп'ютерний тест (40 випадкових коротких запитань з 4 варіантами відповідей, з яких 1 вірна – 1,25 бали за кожну правильну відповідь) або залік (2 теоретичні розгорнутих питання в кожному білеті – максимум 25 балів за правильну відповідь на питання).

Змістовий модуль 2 (Лб) – максимум 50 балів: 16 лабораторних робіт за варіантами (максимум 3 бали за кожну виконану та здану лабораторну роботу) та індивідуальне розрахункове завдання (максимум 2 бали за виконане та захищене розрахункове завдання).

Загалом – максимум 100 балів.

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність.

Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

Дата погодження, підпис
29.08.2023

Завідувач кафедри
Олексій ВОДКА

Дата погодження, підпис
29.08.2023

Гарант ОП
Геннадій ЛЬВОВ