

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»



Ступінь освіти	бакалавр
Спеціальність	122 Комп'ютерні науки
Освітня програма	Комп'ютерні науки
Загальний обсяг	5 кредитів ЄКТС (150 годин)
Тривалість викладання	3 семестр (5,6 чверть)
Заняття:	
лекції:	2 години
лабораторні заняття:	2 години
Мова викладання	українська

Сторінка курсу в СДО НТУ «ДП»: <https://do.nmu.org.ua/course/view.php?id=3214>

Кафедра, що викладає Програмного забезпечення комп'ютерних систем



Викладач:

Мороз Борис Іванович
Професор, д.т.н.

Персональна сторінка

<https://pzks.nmu.org.ua/ua/teachers/morozbi.php>

E-mail: moroz.b.i@nmu.one

1. Анотація до курсу

Інтелектуальний аналіз даних - це дисципліна для вивчення процесу аналізу даних, яка заснована на пошуку в даних схованих закономірностей (шаблонів інформації). Інтелектуальний аналіз даних є мультидисциплінарною областю, яка виникла і розвивається на базі таких наук як математична статистика, розпізнавання образів, штучний інтелект. Вивчення дисципліни «Інтелектуальний аналіз даних» переслідує наступні основні цілі: 1) освоєння теоретичних основ прикладного інтелектуального статистичного аналізу даних; 2) формування практичних навичок для вирішення прикладних завдань з використанням програмного забезпечення інтелектуального аналізу даних. Теоретичний матеріал курсу «інтелектуальний аналіз даних» ґрунтується на припущенні, що дані є різнотипними: безперервними, ранжованими і номінальними. В рамках лекційного курсу розглядаються завдання, алгоритми і

методи інтелектуального даних, що включають розвідувальний аналіз даних, кластеризації, класифікацію (методи машинного навчання), прогнозування, аналіз асоціацій і послідовностей.

1. Мета та завдання курсу

Мета дисципліни – формування стійких знань та засвоєння базових понять сучасної теорії аналізу даних Data Mining. Вивчення загальних принципів та методів застосування сучасної теорії аналізу даних для проектування та розробки систем підтримки прийняття рішень.

Завдання курсу:

У результаті вивчення курсу студенти повинні вивчити: основні завдання і принципи інтелектуального аналізу; розвідувальний (попередній) аналіз даних; аналіз залежності ознак; регресійний аналіз даних; методи машинного навчання: метод опорних векторів, дерева класифікації, ансамблі алгоритмів класифікації, аналіз асоціацій; аналіз послідовностей; методи інтелектуального аналізу для великих обсягів даних; структуру та функціональні можливості програмного забезпечення прикладного інтелектуального аналізу даних.

2. Результати навчання

Дисциплінарні результати навчання сформовано на основі ПРН освітньо-професійної програми «Комп'ютерні науки» першого (бакалаврського) рівня вищої освіти (ПР1, ПР8, ПР12):

- застосовувати знання форм і методів аналізу, обробки і синтезу інформації для розробки систем аналізу даних;
- застосовувати методи і алгоритми інтелектуального аналізу даних на основі технологій Data Mining, Text Mining, Web Mining;
- застосовувати алгоритми інтелектуального аналізу даних для задач аналізу та прогнозування процесів.

3. Структура курсу

Види та тематика навчальних занять	Внесок в загальну оцінку, %
ЛЕКЦІЇ	40
Тема 1. Базові поняття та основні завдання систем аналізу даних Загальні положення. Завдання інтелектуального аналізу даних. Архітектура систем інтелектуального аналізу даних. Приклади застосування інтелектуального аналізу даних. Інтелектуальний аналіз даних в СУБД на прикладі Microsoft SQL Server	
Тема 2. Алгоритми інтелектуального аналізу даних Спрощений алгоритм Байеса, Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором і корекцією моделей. Методи первісної обробки даних. Інструментальні засоби Data Mining, Text Mining, Web Mining. Методи дослідження структури даних: візуалізація даних. Методи опорних векторів, «найближчого сусіда», Байеса	

Види та тематика навчальних занять	Внесок в загальну оцінку, %
<p>Тема 3. Алгоритми інтелектуального аналізу даних Дерева рішень. Постановка задачі класифікації та представлення результатів. Методи побудови правил класифікації. Методи побудови дерев рішень. Методи побудови математичних функцій.. Аналіз багатомірних угруповань. Класифікація об'єктів у випадку невідомих розподілень даних. Методи оцінювання помилок класифікації.</p>	20
<p>Тема 4. Алгоритми інтелектуального аналізу даних Лінійна регресія. Вирішення задачі регресії Сутність задачі прогнозування. Методи вирішення задачі регресії.</p>	
<p>Тема 5. Алгоритми інтелектуального аналізу даних Аналіз часових рядів. Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів</p>	
<p><i>Тестова контрольна робота №1 (за темами 1-5).</i></p>	
<p>Тема 6. Алгоритми інтелектуального аналізу даних Кластеризація. Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.</p>	
<p>Тема 7. Алгоритми інтелектуального аналізу даних Алгоритм Взаємозв'язків. Асоціаційні правила. Сіквенціальний аналіз. Різновиди задач пошуку асоціаційних правил. Метод Apriori. Різновиди методу Apriori.</p>	
<p>Тема 8. Алгоритми інтелектуального аналізу даних Нейронні мережі. Визначення та характеристика нейронних мереж. Область їх застосування. Класифікація нейромережових архітектур. Моделі нейронних мереж. Класифікація методів навчання. Особливості сучасних нейронних мереж. Специфічні галузеві рішення Data Mining на практиці. Бізнес-завдання: в умовах жорсткої конкуренції. Застосування Data Mining при дослідженні задач: Роздрібна торгівля, Банківська справа. Телекомунікації. Страхування. Інші застосування в бізнесі.</p>	
<p>Тема 9. Алгоритми інтелектуального аналізу даних Логістична регресія Статистичні методи обробки даних. Оцінювання параметрів розподілу перевірка статистичних гіпотез; дисперсійний та регресійний аналіз; аналіз часових рядів. Класична модель ARIMA (АРПСС). Експоненційне згладжування з сезонними компонентами. Спектральне розкладання Фур'є. Поліномінальний і регресійний аналіз лагів.</p>	
<p>Тема 10. Базові поняття мови DMX, типи даних і вмісту Послідовне відображення шаблонів даних. Побудова FP-дерев пошуку шаблонів даних. Побудова hash-дерев. Розробка OLAP-кубів під час аналізу багатомірних даних у великих БД. Способи та методи візуального відображення даних.</p>	
<p>Тема 11. Створення структури інтелектуального аналізу Способи інтеграції нечітких та нейронних систем. Нечіткі нейрони. Навчання методами спуска. Нечіткі схеми обміркування. Нейронечіткі класифікатори.</p>	
<p><i>Тестова екзаменаційна робота (за темами 6-11).</i></p>	20
ЛАБОРАТОРНІ ЗАНЯТТЯ	60
<p>Лабораторна робота 1 Використання інструментів «analyze key influencers» і «detect</p>	

Види та тематика навчальних занять	Внесок в загальну оцінку, %
categories»	
<i>Звіт з роботи № 1 та захист лабораторної роботи.</i>	7
Лабораторна робота 2 Використання інструментів «fill from example» і «forecast».	
<i>Звіт з роботи № 2 та захист лабораторної роботи.</i>	7
Лабораторна робота 3 Використання інструментів «highlight exceptions» і «scenario analysis».	
<i>Звіт з роботи № 3 та захист лабораторної роботи.</i>	7
Лабораторна робота 4 Використання інструментів «prediction calculator» і «shopping basket analysis»	
<i>Звіт з роботи № 4 та захист лабораторної роботи.</i>	7
Лабораторна робота 5 Використання інструментів data mining client для excel для підготовки даних	
<i>Звіт з роботи № 5 та захист лабораторної роботи.</i>	8
Лабораторна робота 6 Використання інструментів data mining client для excel для створення моделі інтелектуального аналізу даних	
<i>Звіт з роботи № 6 та захист лабораторної роботи.</i>	8
Лабораторна робота 7 Аналіз точності прогнозу та використання моделі інтелектуального аналізу	
<i>Звіт з роботи № 7 та захист лабораторної роботи.</i>	8
Лабораторна робота 8 Побудова моделі кластеризації, трасування і перехресної перевірки	
<i>Звіт з роботи № 8 та захист лабораторної роботи.</i>	8
РАЗОМ	100

4. Технічне обладнання та/або програмне забезпечення

Використовуються лабораторії кафедри програмного забезпечення комп'ютерних систем (комп'ютерне та мультимедійне обладнання). Дистанційна платформа Moodle, MS Office 365, Microsoft Teams, Data Analyzer.

6. Система оцінювання та вимоги

6.1. Навчальні досягнення здобувачів вищої освіти за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:

Рейтингова шкала	Інституційна шкала
90 – 100	відмінно
74-89	добре
60-73	задовільно
0-59	незадовільно

6.2. Здобувачі вищої освіти можуть отримати підсумкову оцінку з навчальної дисципліни **на підставі поточного оцінювання знань** за умови, якщо набрана кількість балів з поточного тестування та виконання і захисту лабораторних робіт складатиме не менше 60 балів.

Теоретична частина оцінюється за результатами здачі двох контрольних тестових робіт, кожна з яких містить тестові закриті запитання з однією вірною відповіддю (максимальна кількість – 20 балів за кожною тестовою роботою). Загалом за дві контрольні тестові роботи отримується **максимум 40 балів**, тобто 40% від оцінки за дисципліну.

Лабораторні роботи (вісім робіт – у вигляді індивідуального завдання з кожною, розподіл % див. в таблиці розділу 4) виконуються у письмовому вигляді (звіт з кожної роботи оцінюється в межах балів, представлених в таблиці розділу 4, загалом лабораторні враховуються як 60% (максимум 60 балів). У сумі за лабораторну частину курсу при поточному оцінюванні отримується **максимум 60 балів**.

Отримані бали за теоретичну частину та лабораторні роботи додаються і є поточною успішністю за вивчення навчальної дисципліни. Максимально за поточною успішністю здобувач вищої освіти може набрати 100 балів.

Максимальне оцінювання поточного контролю в балах:

Теоретична частина	Практична частина	Разом
40	60	100

6.3. Критерії оцінювання підсумкової роботи. У випадку якщо здобувач вищої освіти за поточною успішністю отримав менше 60 балів та/або прагне поліпшити оцінку, проводиться **підсумкове оцінювання (екзамен)** під час сесії.

Екзамен проводиться у вигляді комплексної контрольної роботи, яка включає запитання з теоретичної та практичної частини курсу. Білет складається з **40 тестових завдань** з чотирма варіантами відповідей, одна правильна відповідь оцінюється в 2 бали (**разом 80 балів**) та **2 завдань** з практичної частини, кожне з запитань оцінюється максимум у 10 балів (**разом 20 балів**), причому:

- 10 балів – відповідність еталону;
- 8 балів – відповідність еталону з незначними помилками;
- 6 балів – часткова відповідність еталону, питання повністю не розкриті;
- 4 бали – невідповідність еталону, але відповідність темі запитання;
- 0 балів – відповідь не наведена або не відноситься до теми запитання.

Отримані бали за тестові завдання та завдання з практичної частини додаються і є підсумковою оцінкою за вивчення навчальної дисципліни. Максимально за підсумковою роботою здобувач вищої освіти може набрати 100 балів.

7. Політика курсу

7.1. Політика щодо академічної доброчесності. Академічна доброчесність здобувачів вищої освіти є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадкування даних чи фактів, що використовуються в освітньому процесі). Політика щодо академічної доброчесності регламентується положенням "Положення про систему запобігання та виявлення плагіату у Національному технічному університеті "Дніпровська політехніка" <https://cutt.ly/MCfh5kv>

У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

7.2. Комунікаційна політика. Здобувачі вищої освіти повинні мати активовану корпоративну університетську пошту.

Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

7.3. Політика щодо перескладання. Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання підсумкового оцінювання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

7.4 Політика щодо оскарження оцінювання. Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку.

8. Рекомендовані джерела інформації

Базова:

1. Марченко О. О., Россада Т. В. Актуальні проблеми Data Mining: навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. Київ: Київський національний університет імені Тараса Шевченка, 2017. 150 с.

2. Top 33 Data Mining software. URL: <https://www.predictiveanalyticstoday.com/top-data-mining-software>

3. Tierney B. Predictive Analytics Using Oracle Data Miner: Develop & Use Data Mining Models in Oracle Data Miner, SQL & PL/SQL: Oracle Press, 2014. 429 p.

4. Oracle Advanced Analytics Customer Success Stories. URL: <https://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odm-customers-086483.html>

Додаткова:

1. Geourjon A.-M., Laporte B., Coundoul O., Gadiaga M. Inspecting less to inspect better: The use of data mining for risk management by customs administrations measurement Applied to Customs and Tax Administrations in Developing Countries ed. by T. Cantens, R. Ireland and G. Raballand. Washington, DC: World Bank, 2012. November.

2. 50 top free data mining software. URL: <https://www.predictiveanalyticstoday.com/top-free-data-mining-software/#>