



Технології інтелектуального аналізу даних

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>122 Комп'ютерні науки</i>
Освітня програма	<i>Комп'ютерні науки</i>
Статус дисципліни	<i>Обов'язкова</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>1 курс, весняний семестр</i>
Обсяг дисципліни	<i>150 годин/ 5 кредитів ЕКТС (лекції – 36 год., лабораторні роботи – 18 год., СРС – 96 год.)</i>
Семестровий контроль/ контрольні заходи	<i>Екзамен, МКР</i>
Розклад занять	<i>https://schedule.kpi.ua/ (2 год. лекційних та 1 год. лабораторних робіт на тиждень)</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: доктор технічних наук, професор Мухін Вадим Євгенійович, v_mukhin@i.ua, (067)5087684 Лабораторні: доктор технічних наук, професор Мухін Вадим Євгенійович, v_mukhin@i.ua, (067)5087684</i>
Розміщення курсу	<i>Платформа дистанційного навчання "Сікорський"</i>

Програма навчальної дисципліни

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

Цілі дисципліни: сформувати навички щодо дослідження і виявлення алгоритмами, засобами штучного інтелекту в «сирих даних» прихованих структур, шаблонів (паттернів) або залежностей, які раніше не були відомі, нетривіальні, практично корисні та доступні для інтерпретації людиною і необхідні для ухвалення рішень в різних сферах діяльності.

Завданням дисципліни «Технології інтелектуального аналізу даних» є надання студентам ґрунтовних знань в області аналітичних досліджень інформаційного простору, вивчення методів створення, добування, консолідації, переробки, трансформації та аналізу даних.

Предметом вивчення дисципліни є основні положення, методи та алгоритми інтелектуального аналізу даних, зокрема із залученням машинного навчання та штучного інтелекту, та їх комп'ютерна реалізація з використанням мови програмування Python, зокрема спеціальних бібліотек, таких як Pandas, Numpy, sklearn, Keras.

Компетентності:

ЗК 1 Здатність до абстрактного мислення, аналізу та синтезу.

ЗК 2 Здатність застосовувати знання у практичних ситуаціях.

ЗК 5 Здатність вчитися й оволодівати сучасними знаннями.

ЗК 7 Здатність генерувати нові ідеї (креативність).

ФК 1 Усвідомлення теоретичних засад комп'ютерних наук.

ФК 5 Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.

ФК 7 Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

ФК 17 Здатність вибирати адекватні методи і технології обчислювального інтелекту та машинного навчання, включаючи методи глибокого навчання, еволюційного моделювання, генетичні алгоритми, та використовувати їх для вирішення задач прогнозування, керування, прийняття рішень, класифікації та інтелектуального аналізу даних в умовах невизначеності та неповної інформації.

Програмні результати навчання:

ПРН 1 Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.

ПРН 2 Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур.

ПРН 6 Розробляти концептуальну модель інформаційної або комп'ютерної системи.

ПРН 7 Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

ПРН 9 Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

ПРН 11 Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

ПРН 12 Проектувати та супроводжувати бази даних та знань.

ПРН 14 Тестувати програмне забезпечення.

ПРН 18 Збирати, формалізувати, систематизувати і аналізувати потреби та вимоги до інформаційної або комп'ютерної системи, що розробляється, експлуатується чи супроводжується.

ПРН 19 Аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій.

ПРН 20 Створювати та досліджувати інформаційні та математичні моделі систем і процесів, що досліджуються, зокрема об'єктів автоматизації.

ПРН 26 Застосовувати технології обчислювального інтелекту та інтелектуального аналізу даних, зокрема, нейронні мережі, нечіткі нейронні мережі, нейронні мережі глибокого навчання, методи машинного навчання для проектування та адаптації інтелектуальних систем прийняття рішень в різних предметних сферах.

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

Курс «Технології інтелектуального аналізу даних» є одним із основних курсів підготовки магістрів науковців ОНП «Комп'ютерні науки».

Цей курс підсумовує раніше прочитані спеціальні дисципліни в напрямку теорії та систем прийняття рішень і дає систематизоване детальне викладання основ теорії, методів та засобів побудови систем обчислювального інтелекту та їх застосування в системах прийняття рішень в економіці, бізнесі та фінансовій сфері. Матеріали курсу широко використовуються для підготовки магістерської дисертації.

3. Зміст навчальної дисципліни

Освітній компонент охоплює такі теми:

Тема 1. Основні задачі Data Mining

Тема 2. Інформативність факторів. Способи підвищення інформативності

Тема 3. Алгоритм Фаррара-Глобера

Тема 4. Формування множини значимих факторів

Тема 5. Підходи до машинного навчання з вчителем та без нього

Тема 6. Навчання нейронних мереж прямого поширення.

Тема 7. Послідовні шаблони

Тема 8. Методи класифікації

Тема 9. Кластеризація. Гіпотеза компактності. Еволюційний пошук

Тема 10. Деревя рішень для класифікації та регресії

Тема 11. Класифікація з незбалансованими даними

Тема 12. Байєсові мережі

Тема 13. Відновлення інформації в наборах даних

4. Навчальні матеріали та ресурси

1. Базова література

1. Zgurovsky M., Zaychenko Yu. Big Data: Conceptual Analysis and Applications. Springer Nature Switzerland AG. – 2020. – 275 p. (НТБ ім. Г.І. Денисенка)
2. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних : навч. посіб. / В. О. Гороховатський, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 92 с. – <https://openarchive.nure.ua/handle/document/15868>
3. Інтелектуальний аналіз даних : практикум / М. Т. Фісун, І. О. Кравець, П. П. Казмірчук, С. Г. Ніколенко; М-во освіти і науки України, Чорноморський державний університет ім. Петра Могили. – Львів: Новий Світ–2000, 2023. – 160 с. (НТБ ім. Г.І. Денисенка)
4. Москаленко В. В. Моделі і методи інтелектуального аналізу багатовимірних даних за умов апіорної невизначеності. – Суми : Сумський державний університет, 2020. – 184 с. – https://essuir.sumdu.edu.ua/bitstream-download/123456789/77692/3/Moskalenko_analiz_danykh.pdf

2. Допоміжна література

1. M. Zgurovsky, Yu. Zaychenko. Fundamentals of computational intelligence- System approach. Springer, 2016.-275 p. (НТБ ім. Г.І. Денисенка)
2. Yuriy Zaychenko. Problem of Fuzzy portfolio optimization and its solution with application of forecasting methods. Scholar Press. – 2015, 54 p. url: <http://journal.iasa.kpi.ua/article/view/65695>
3. Huyen C. Designing Machine Learning Systems. - O'Reilly Media. USA – 2022. – 463 p. (за запитом викладачу)
4. Kantarzic M. Data Mining. Concepts, Models, Methods and Algorithms / M. Kantarzic, 3rd Ed. – Publisher : Wiley, 2019. – 672 p (за запитом викладачу)
5. Землянський О.М., Пашинська Н.М., Снитюк В.Є. Інформаційно-аналітичні технології прогнозування наслідків хімічних аварій, Київ, ВПЦ “Київський університет”, 2017, 167 с. (за запитом викладачу)
6. Ланде Д.В. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник / Д. В. Ланде, І. Ю. Субач, Ю. Є. Бояринова ; КПІ ім. Ігоря Сікорського. — Київ : КПІ ім. Ігоря Сікорського, 2018. – 300 с. (НТБ ім. Г.І. Денисенка)
7. Черняк О.І. Інтелектуальний аналіз даних: Підручник / Черняк О.І., Захарченко П.В./ К.: Знання, 2014 р. - 599 с. url: https://moodle.znu.edu.ua/pluginfile.php/767346/mod_resource/content/1/%D0%9F%D1%96%D0%B4%D1%80%D1%83%D1%87%D0%BD%D0%B8%D0%BA.pdf

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Навчальна дисципліна охоплює 36 годин лекцій та 18 годин лабораторних робіт, а також виконання модульної контрольної роботи.

У межах дисципліни заплановано такі види занять:

- лекції;
- лабораторні роботи;
- самостійна робота.

Теми дисципліни взаємопов'язані, матеріал вивчається в логічній послідовності. На лекціях розкриваються найсуттєвіші теоретичні питання, які дають змогу забезпечити студентам можливість глибокого самостійного вивчення всього програмного матеріалу. Теми та порядок виконання практичних завдань сформовано в логічній послідовності і повністю узгоджується і підтримується лекційним матеріалом. Теоретичні та практичні знання та навички поглиблюються шляхом самостійної роботи з використанням рекомендованої літератури та веб-джерел.

**Термін
виконання
(тиждень)**

Назви розділів і тем

Тема 1. Основні задачі Data Mining	
1	Лекція 1. Задачі Data Mining Лекція 2. Ключові відмінності між засобами Data Mining і OLAP. Концепції сховищ даних
Тема 2. Інформативність факторів	
2	Лекція 3. Ентропія та інформативність факторів
3	Лекція 4. Способи підвищення інформативності Лабораторна робота 1. Препроцесинг інформації та побудова множинної лінійної регресії
Тема 3. Алгоритм Фаррара-Глобера	
4	Лекція 5. Алгоритм Фаррара-Глобера перевірки вхідних даних
Тема 4. Множина значимих факторів	
5	Лекція 6. Формування множини значимих факторів Лабораторна робота 2. Визначення множини значимих факторів та побудова множинної нелінійної регресії
Тема 5. Підходи до машинного навчання з вчителем та без нього	
6	Лекція 7. Підходи до машинного навчання з вчителем
7	Лекція 8. Підходи до машинного навчання без вчителя
Тема 6. Навчання нейронних мереж прямого поширення	
7	Лекція 9. Методи навчання нейронних мереж прямого поширення
Тема 7. Послідовні шаблони	
8	Лекція 10. Пошук послідовних шаблонів
9	Лабораторна робота 3. Пошук асоціативних правил та послідовних шаблонів
Тема 8. Методи класифікації	
9	Лекція 11. Класичні, ітераційні та евристичні методи класифікації
10	Лабораторна робота 4. Класифікація об'єктів у змішаному просторі. Частина 1
11	Лабораторна робота 5. Класифікація об'єктів у змішаному просторі. Частина 2
Тема 9. Кластеризація. Гіпотеза компактності. Еволюційний пошук	
11	Лекція 12. Кластеризація
12	Лекція 13. Гіпотеза компактності. Еволюційний пошук

Термін виконання (тиждень)	Назви розділів і тем
13	Лабораторна робота 6. Кластеризація на множині об'єктів дійсного простору
Тема 10. Дерева рішень для класифікації та регресії	
13	Лекція 14. Алгоритми побудови дерев рішень для класифікації та регресії
Тема 11. Класифікація з незбалансованими даними	
14	Лекція 15. Підходи та методи до класифікації з незбалансованими даними
15	Лекція 16. Активне навчання. Метрики для активного навчання Лабораторна робота 7. Добування знань з даних, порівняння алгоритмів
Тема 12. Байєсові мережі	
16	Лекція 17. Байєсові мережі
Тема 13. Відновлення інформації в наборах даних	
17	Лекція 18. Підходи та методи до відновлення інформації в наборах даних Лабораторна робота 8. Створення бази даних з MySQL та побудова аналітичних репортів в Tableau. Частина 1
18	Лабораторна робота 9. Створення бази даних з MySQL та побудова аналітичних репортів в Tableau. Частина 2 Модульна контрольна робота

ДИДАКТИЧНІ МЕТОДИ:

На лекційних заняттях: Лекція, пояснення, мозковий штурм, проблемні завдання.

На практичних заняттях: Завдання до виконання, опитування та тестування студентів

6. Самостійна робота здобувача вищої освіти

Важливим аспектом для якісного засвоєння матеріалу, відпрацювання прийомів і алгоритмів вирішення основних завдань дисципліни є самостійна робота, що дає змогу перетворити отримані знання в об'єкт власної діяльності. Самостійна робота включає в себе читання літератури, огляд літератури за темою, виконання звітів з лабораторних робіт, підготовку до їх захисту та підготовку до іспиту.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Викладач повинен звернути увагу студентів на те, що дисципліна займається розробкою і застосуванням методів та технологій ШІ в прикладних задачах розпізнавання образів, класифікації, кластерного аналізу в різних областях людської діяльності в умовах неповноти та невизначеності.

Рекомендовані методи навчання: проектний метод, імітаційні вправи, презентація та опитування студентів.

Студенту рекомендується вести докладний конспект лекцій і фіксувати основні результати практичних занять.

Відсутність на аудиторному занятті не передбачає нарахування штрафних балів, оскільки фінальний рейтинговий бал студента формується виключно на основі оцінювання результатів навчання. Разом з тим, захист лабораторних робіт оцінюватиметься під час аудиторних занять.

Кожен студент має право відпрацювати пропущені з поважної причини (лікарняний, мобільність тощо) заняття за рахунок самостійної роботи. Детальніше за посиланням: <https://kpi.ua/files/n3277.pdf>.

Студент може підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами. Студенти мають право аргументовано оскаржити результати контрольних заходів, пояснивши з яким критерієм не погоджуються відповідно до оціночного.

Календарний контроль проводиться з метою підвищення якості навчання студентів та моніторингу виконання студентом вимог силабусу.

Критерій		Перший календарний контроль	Другий календарний контроль
Термін календарного контролю		Тиждень 8	Тиждень 14
Умови отримання позитивної оцінки	Поточний рейтинг	≥ 10 балів	≥ 20 балів

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Засвоєння знань та умінь в ході вивчення дисципліни може бути доступним для більшості осіб з особливими освітніми потребами, окрім здобувачів з серйозними вадами зору, які не дозволяють виконувати завдання за допомогою персональних комп'ютерів, ноутбуків та/або інших технічних засобів.

У ході виконання завдань студентам може бути рекомендовано звернутися до англomовних джерел.

Відповідно до Положення про систему оцінювання результатів навчання сума всіх заохочувальних балів не може перевищувати 10% рейтингової шкали оцінювання. Штрафні бали призначаються студентам, які несвоєчасно захистили лабораторні роботи, не виконали роботи в повному обсязі, не виконали вимоги щодо якісного оформлення результатів робіт. Відповідно до Положення про систему оцінювання результатів навчання сума всіх штрафних балів не може перевищувати 10% рейтингової шкали оцінювання. Зокрема, здача лабораторної роботи пізніше встановленого терміну – -1 бал (за кожен тиждень пізніше встановленого терміну); участь у модернізації лабораторних робіт, удосконалення дидактичних матеріалів з дисципліни надається +5-10 заохочувальних балів.

Підготування до лабораторних робіт та контрольних заходів здійснюється під час самостійної роботи студентів з можливістю консультування з викладачем у визначений час консультацій або за допомогою електронного листування (електронна пошта, месенджери).

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

Поточний контроль: фронтальний (усний, письмовий), лабораторні роботи.

Календарний контроль: проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Рейтингова система оцінювання охоплює всі види тестування: контрольні роботи, активність на лабораторних заняттях та якість захисту лабораторних робіт. Кожен студент отримує свій підсумковий рейтинг за дисципліною.

Рейтинг студента з кредитного модуля у семестрі складається з балів, які він отримує за:

- написання модульної контрольної роботи;
- робота на лабораторних заняттях та захист лабораторних робіт;
- відвідування лекцій та написання конспекту під час лекцій;
- відповіді на екзамені.

Рейтингова оцінка r_p з дисципліни формується як сума балів стартової оцінки r_c з ваговим коефіцієнтом 0,6 та екзаменаційної оцінки r_e з ваговим коефіцієнтом 0,4:

$$r_p = 0,6 r_c + 0,4 r_e$$

Стартова система оцінювання r_c , максимальна величина якої становить 60 балів, складається з двох оцінок – оцінки за виконання лабораторного практикуму $r_{\text{ЛАБ}}$ (максимальна величина якої становить 42 бали) та оцінки з модульної контрольної роботи $r_{\text{МКР}}$ (максимальна величина якої становить 18 балів):

$$r_c = r_{\text{ЛАБ}} + r_{\text{МКР}}$$

Система рейтингових (вагових) балів та критеріїв оцінювання:

Метод оцінювання	Кількість	Мінімальна оцінка в балах	Максимальна оцінка в балах
Лабораторні роботи	7	3	6
Модульна контрольна робота	1	0	18
Стартовий рейтинг		36	60
Іспит	1		40
Підсумковий рейтинг		60	100

Модульна контрольна робота містить 18 комплексних питань тестового, розрахункового або відкритого (питання, яке вимагає розгорнутої текстової відповіді) типу, які оцінюються в один бал. За правильну відповідь на питання студент отримує 1 бал, неправильну – 0 балів.

Кожна з семи виконаних та захищених лабораторних робіт оцінюється за 6-бальною шкалою: максимальна оцінка – 6, мінімальна позитивна оцінка – 3. Таким чином, сумарна максимальна оцінка за виконання всього циклу лабораторних робіт складає 42 бали, а сумарна мінімальна позитивна оцінка – 21 бал.

Стартова система оцінювання (максимум 60 балів) призначена для оцінювання якості і своєчасності виконання циклу лабораторних занять впродовж навчального семестру, а також якості і своєчасності виконання завдань модульної контрольної роботи.

Екзаменаційну систему оцінювання (максимум 40 балів) побудовано на основі повноти відповідей студента на екзаменаційні запитання, а також точності та аргументованості відповідей на додаткові запитання екзаменатора. Під час оцінки знань студентів на екзамені враховується розуміння питань та повнота відповідей на них, розуміння студентом зв'язку поставлених запитань з іншими темами дисципліни.

Умови допуску до екзамену: захищені усі лабораторні роботи та мінімальна позитивна стартова оцінка (сумарна оцінка з лабораторного практикуму та модульної контрольної роботи) 36 балів.

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
<36 або не виконано інші умови допуску до екзамену	Не допущено

Робочу програму навчальної дисципліни (силабус):

Складено завідуючим кафедри системного проектування, доктором технічних наук, професором **Мухін Вадим Євгенійович**

Ухвалено кафедрою системного проектування (протокол № 10 від 12 червня 2023 р.)

Погоджено Методичною комісією ІПСА (протокол № 4 від 16 червня 2023 р.)

Погоджено науково-методичною комісією КПІ ім. Ігоря Сікорського зі спеціальності 122 (протокол № 6 від 27 червня 2023 р.)