



Інтелектуальний аналіз даних в бізнес-аналітиці

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>Системи і методи штучного інтелекту</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>очна (денна)</i>
Рік підготовки, семестр	<i>I курс, весняний семестр</i>
Обсяг дисципліни	<i>150 годин (5 кредитів) лекційних занять: 36 год практичних занять: 18 год, самостійна робота студента: 96 год.</i>
Семестровий контроль/ контрольні заходи	<i>Екзамен, модульна контрольна робота</i>
Розклад занять	<i>https://schedule.kpi.ua/ 2 год лекційних та 1 год практичних занять на тиждень</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: к.ф.-м.н. Лазаренко Ірина Сергіївна, email: irynalazar@gmail.com Практичні: к.ф.-м.н. Лазаренко Ірина Сергіївна, email: irynalazar@gmail.com</i>
Розміщення курсу	<i>Сервісу Zoom / Google Meet (за узгодженням зі студентами)</i>

Програма навчальної дисципліни

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

Метою навчальної дисципліни є формування у студентів компетентностей:

ЗДАТНОСТЕЙ:

- застосовувати комп'ютерні технології та програмне забезпечення з обробки даних для вирішення бізнес завдань, аналізу інформації та підготовки аналітичних звітів;
- обробляти та аналізувати неструктуровані великі масиви даних для вирішення бізнес задач;

Після засвоєння навчальної дисципліни студенти мають продемонструвати такі результати навчання:

ЗНАННЯ:

- методів машинного навчання та багатовимірного статистичного аналізу даних;
- технологій використання сучасних програмних засобів інтелектуального аналізу даних.
- критеріїв порівняння моделей і методів інтелектуального аналізу даних

УМІННЯ:

- використовувати методи інтелектуального аналізу даних для вирішення практичних задач бізнес-аналітики
- застосовувати сучасні програмні засоби інтелектуального аналізу даних.

- обирати оптимальні методи та моделі інтелектуального аналізу даних при вирішенні конкретних задач управління та бізнесу.

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

Курс ґрунтується на засвоєних студентами дисципліни: «Оброблення надвеликих масивів даних» та «Методи та технології обчислювального інтелекту». Курс передує практиці та проектуванню магістерської дисертації.

3. Зміст навчальної дисципліни

Тема 1. Вступ до інтелектуального аналізу даних.

Тема 2. Основні теоретичні засади лінійної алгебри та теорії функції.

Тема 3. Основні методи візуалізації даних.

Тема 4. Етапи підготовки даних. Препроцесінг.

Тема 5. Лінійна та нелінійна регресії.

Тема 6. Часові ряди

Тема 7. Основні засади навчання на розмічених даних (з вчителем).

Тема 8. Методи класифікації: kNN (*k-nearest neighbors algorithm*); DT (*Decision tree*).

Тема 9. Методи класифікації: LR (*Logistic regression*), SVM (*support vector machine*).

Тема 10. Пошук структури в даних. Навчання без вчителя

Тема 11. Методи кластеризації: K-means, EM-алгоритм.

Тема 12. Методи кластеризації: Ієрархічна кластеризація, DBSCAN (*Density-based spatial clustering of applications with noise*)

Тема 13. Метод головних компонент. Матричний розклад.

Тема 14. L1 та L2 регуляризація.

Тема 15. Рекомендовані системи в машинному навчанні.

Тема 16. Оцінки якості моделі. АВ-тестування

Тема 17. Бізнес аналіз та рекомендації рішень із застосуванням технологій Data Science та Machine Learning.

Тема 18. Data Driven підхід для розвитку бізнесу

4. Навчальні матеріали та ресурси

1. Базова література

1. Черняк О.І. Інтелектуальний аналіз даних: Підручник / О.І. Черняк, П.В. Захарченко ; Київ. нац. ун-т ім. Т. Шевченка. — К.: Знання, 2014. — 599 с.
https://moodle.znu.edu.ua/pluginfile.php/767346/mod_resource/content/1/Підручник.pdf
2. Олійник А. О. Інтелектуальний аналіз даних : Навчальний посібник / А. О. Олійник, О. О. Олійник, С. О. Субботін. — Запоріжжя : ЗНТУ, 2012. — 278 с. (НТБ ім. Г.І. Денисенка)
3. Том Фоусет, Фостер Провост «Data Science для бізнесу. Як збирати, аналізувати і використовувати дані», - Наш формат, Київ, 2019. — 400 с. (НТБ ім. Г.І. Денисенка)
4. Інтелектуальний аналіз даних: практикум [Комплект] / Фісун М.Т., Кравець І.О., Казмірчук П.П., Ніколенко С.Г. — Л.: «Новий Світ – 2000», 2020. — 162 с. (НТБ ім. Г.І. Денисенка)

2. Допоміжна література

1. Oswald Campesato, Python 3 and Data Analytics Pocket Primer. Mercury Learning and Information, 2021. (НТБ ім. Г.І. Денисенка)
2. Davy Cielen, Arno Meysman, and M. Ali, Introducing data science : big data, machine learning, and more, using Python tools. Shelter Island, Ny: Manning Publications, 2016. (за запитом викладачу)
3. Копей В.Б. Мова програмування Python для інженерів і науковців: Навчальний посібник. Івано-Франківськ : ІФНТУНГ, 2019. 274 с. (НТБ ім. Г.І. Денисенка)
4. Висоцька В.А., Оборська О.В. Python: алгоритмізація та програмування: навчальний посібник – Львів:Видавництво «Новий Світ – 2000», 2021. — 514 с. (НТБ ім. Г.І. Денисенка)

5. Методика опанування навчальної дисципліни (освітнього компонента)

В рамках дисципліни заплановано наступні види навчальних занять:

- лекції;
- практичні заняття (комп'ютерні практикуми);
- самостійна робота.

Теми дисципліни взаємозв'язані, матеріал вивчається в логічній послідовності. На лекціях розкриваються найбільш суттєві теоретичні питання, які дозволяють забезпечити студентам можливість глибокого самостійного вивчення всього програмного матеріалу. Теми та порядок виконання практичних завдань сформовано в логічній послідовності і повністю узгоджуються з лекційним матеріалом. Теоретичні і практичні знання поглиблюються шляхом самостійної роботи з використанням рекомендованої літератури та глобальної мережі Internet.

На заняттях використовуються звичайна дошка та/або інтерактивна дошка, віртуальна дошка (в умовах карантинних обмежень), а також презентації лекцій з використанням мультимедіа проектора або дистанційно.

Контроль засвоєння навчального матеріалу здійснюється індивідуальним опитуванням, перевіркою комп'ютерних практикумів, індивідуальних завдань, модульною контрольною роботою та семестровим екзаменом.

Практичні заняття

Основні завдання практичних занять:

Навчитись аналізувати дані різної природи, використовуючи алгоритми інтелектуального аналізу даних. Знати й вміти застосовувати відповідні бібліотеки мови програмування Python, будувати моделі, проводити статистичний аналіз результатів моделювання. Застосовувати інструментарій інтелектуального аналізу до бізнес-процесів.

6. Самостійна робота студента/аспіранта

Індивідуальні завдання сприяють поглибленому засвоєнню алгоритмів машинного навчання при побудові та аналізу моделі. Методичні рекомендації до виконання індивідуальних завдань та термін виконання надає лектор всім групам потоку. Індивідуальні завдання виконуються за тематикою, яка вибирається студентом самостійно після обговорення з викладачем і є дотичною або охоплює тему наукових інтересів студента.

- Підготовка до практичних занять (повторення лекції, виконання комп'ютерного практикуму) - 1 год на тиждень;
- Виконання індивідуальних завдань – 15 годин.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Мета проведення практичних занять з навчальної дисципліни полягає в систематизації, закріпленні і поглибленні знань, одержаних при вивченні теоретичного та практичного курсу дисципліни; у набутті студентами професійних знань і вмінь, необхідних для практичної роботи у сфері моделювання та аналітики; вихованні загальної інформаційної культури, математичної обізнаності для роботи на ринку інформаційних технологій, обробки та аналізу даних за рахунок уявлень просторових взаємозв'язків, відображень та трансформацій.

Після здачі та перевірки індивідуальних завдань студент захищає її шляхом відповіді на теоретичні та практичні питання.

Додаткові бали можуть бути надані у зв'язку з активною роботою на заняттях.

Штрафні бали можуть бути зняті за несвоєчасну здачу індивідуальних завдань.

Політика перескладань:

- можна перескласти одну контрольну роботу за семестр, при чому максимум балів встановлюється на рівні двох третіх від початкового;
- екзамен перескладається за умови отримання незадовільної оцінки протягом додаткової сесії.

Всі дії студента та викладача мають бути у відповідності до:

[Кодекс честі КПІ ім. ІгоряСікорського](#)

[Положення про систему запобігання академічному плагіату](#)

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

1. Рейтинг студента з кредитного модуля розраховується виходячи із 100-бальної шкали та складається з балів, що він отримує за:

- виконання модульної контрольної роботи (МКР);
- роботи на 8 комп'ютерних практикумах
- виконання індивідуальних завдань (ІЗ).

2. Критерії нарахування балів.

2.1. Модульна контрольна робота складає 10 балів:

- «відмінно» – повний розв'язок (не менше 90% потрібної інформації) – 9-10 балів;
- «добре» – присутній правильний хід розв'язку з помилками в обрахунках (не менше 75% потрібної інформації) – 7-8 балів;
- «задовільно» – частково розв'язані задачі (не менше 60% потрібної інформації) та незначні помилки – 5-6 балів;
- «незадовільно» – відповідь не відповідає вимогам до «задовільно» – менше 5 балів.

2.2. Практичне заняття оцінюються в 5 балів:

- своєчасне та повне виконання комп'ютерного практикуму – 5 балів;
- несвоєчасне або неповне виконання комп'ютерного практикуму – не менше 1 балу;

2.3. Індивідуальні завдання оцінюються в 20 балів за такими критеріями:

- «відмінно» – творчий підхід до розкриття проблеми – 20-18 балів;
- «добре» – глибоке розкриття проблеми, відображена власна позиція – 17-15 балів;
- «задовільно» – обґрунтоване розкриття проблеми з певними недоліками – 12-14 бали;
- «незадовільно» – завдання не виконане, ІЗ не зараховано – менше 12 балів.

Наявність позитивної оцінки з індивідуальних завдань є умовою допуску до екзамену.

2.4. Екзаменаційна контрольна робота оцінюється в 30 балів. Контрольне завдання цієї роботи складається з трьох запитань: двох теоретичних з переліку, що наданий у додатку до робочої програми КМ, та одного практичного.

Кожне запитання оцінюється у 10 балів за такими критеріями:

- «відмінно» – повна відповідь (не менше 90% потрібної інформації), надано повний розв'язок задачі – 10 - 9 балів;
- «добре» – достатньо повна відповідь (не менше 75% потрібної інформації), присутній правильний хід розв'язку з помилками в обрахунках – 8 - 6 балів;
- «задовільно» – неповна відповідь (не менше 60% потрібної інформації, що виконана згідно з вимогами до «стереотипного» рівня та деякі помилки), частково розв'язані задачі – 5 - 4 балів;
- «незадовільно» – незадовільна відповідь – менше 4 балів.

3. Умовою позитивної першої атестації є отримання не менше 20 балів, другої атестації – отримання не менше 45 балів за умови зарахування індивідуальних завдань.

4. Сума рейтингових балів, отриманих студентом протягом семестру, за умови зарахування індивідуальних завдань та екзаменаційної контрольної роботи переводиться до підсумкової оцінки згідно з таблицею (п.5).

6. Таблиця переведення рейтингових балів до оцінок:

Бали	Оцінка
100...95	Відмінно
94...85	Дуже добре
84...75	Добре
74...65	Задовільно
64...60	Достатньо
Менше 60	Незадовільно
Із не зараховано менше 12	Не допущено

9. Додаткова інформація з дисципліни (освітнього компонента)

Теоретичні питання до залікової контрольної роботи

1. Поняття інтелектуального аналізу даних (ІАД).
2. Етапи моделювання
3. Основні об'єкти та операції лінійної алгебри для ІАД
4. Основні об'єкти та операції теорії функції для ІАД
5. Основні методи візуалізації даних.
6. Етапи підготовки даних. Препроцесінг .
7. Лінійна регресія.
8. Нелінійна регресія
9. Часові ряди
10. Основні засади навчання на розмічених даних (з вчителем).
11. Методи класифікації: kNN (k-nearestneighborsalgorithm);
12. Методи класифікації:DT (Decisiontree).
13. Методи класифікації:LR(Logisticregression)
14. Методи класифікації:SVM(support vector machine).
15. Пошук структури в даних. Навчання без вчителя
16. Методи кластеризації: K-means
17. Методи кластеризації: EM-алгоритм.
18. Методи кластеризації: Ієрархічна кластеризація
19. Методи кластеризації: DBSCAN(Density-basedspatialclusteringofapplicationswithnoise)
20. Метод головних компонент.
21. Матричний розклад.
22. L1 та L2 регуляризація.
23. Рекомендовані системи в машинному навчанні: Колаборативна фільтрація
24. Рекомендовані системи в машинному навчанні: заснована на контенті
25. Оцінки якості моделі. АВ-тестування
26. Основи глибинного навчання.

Робочу програму навчальної дисципліни (силабус):

Складено к.ф.-м.н., доцент Лазаренко Ірина Сергіївна

Ухвалено кафедрою ШІ (протокол № 1 від 5.07.2022)

Погоджено Методичною комісією НН ІПСА (протокол № 8 від 17.06.2022)