

Силабус курсу:



СХІДНОУКРАЇНСЬКИЙ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВОЛОДИМИРА ДАЛЯ

Аналіз великих даних або машинне навчання

<i>Ступінь вищої освіти:</i>	доктор філософії
<i>Спеціальність:</i>	122 «Комп'ютерні науки»
<i>Рік підготовки:</i>	2
<i>Семестр викладання:</i>	весняний
<i>Кількість кредитів ЄКТС:</i>	4
<i>Мова(-и) викладання:</i>	українська, англійська
<i>Вид семестрового контролю</i>	іспит

Автор курсу та лектор:

к.ф.-м.н., доцент Хорошун Г.М.

вчений ступінь, вчене звання, прізвище, ім'я та по-батькові

доцент кафедри комп'ютерних наук та інженерії

посада

horoshun@snu.edu.ua

електронна адреса

+380661719306

телефон

Telegram

месенджер

Online

консультації

Викладач лабораторних занять:*

вчений ступінь, вчене звання, прізвище, ім'я та по-батькові

посада

електронна адреса

телефон

месенджер

консультації

Викладач практичних занять:*

к.ф.-м.н., доцент Хорошун Г.М.

вчений ступінь, вчене звання, прізвище, ім'я та по-батькові

доцент

посада

електронна адреса

телефон

месенджер

консультації

* – 1) дані підрозділи вносяться до силабусу в разі, якщо практичні та (або) лабораторні заняття проводить інший викладач, котрий не є автором курсу та лектором; 2) припустимо змінювати назву підрозділу на «Викладач лабораторних та практичних занять:», якщо лабораторні та практичні заняття проводить один викладач, котрий не є автором курсу та лектором.

Анотація навчального курсу

Цілі вивчення курсу:

Формування у здобувачів вищої освіти професійних компетентностей щодо застосування сучасних наукових концепцій, понять, методів та технологій обробки великих даних з використанням методів машинного навчання.

Мета лекційних занять — надати теоретичну базу, яка охоплює ключові поняття і підходи в галузі машинного навчання та аналізу великих даних, зокрема історію розвитку, принципи, галузі застосування, джерела та типи даних, методи і програмні засоби, а також алгоритми машинного навчання та їх застосування у різних сферах.

Мета практичних занять — розвинути у студентів навички вибору і застосування відповідних алгоритмів і методів машинного навчання на практиці, враховуючи характеристики даних, часові та обчислювальні обмеження, а також особливості завдань. Студенти навчатимуться оцінювати ефективність алгоритмів, використовувати лінійні класифікатори, такі як SVM, і освоють процедуру вибору оптимального методу залежно від вимог до результатів.

Мета самостійної роботи — систематизація і закріплення теоретичних знань, а також вдосконалення практичних умінь. Студенти вивчатимуть етапи процесу машинного навчання, технологію CRISP-DM, питання зміщення і розкиду помилок (Bias and Variance), методи контрольованого і неконтрольованого навчання, а також освоють оцінку результатів роботи алгоритмів машинного навчання, що підготує їх до самостійного розв'язання задач аналізу даних у професійному контексті.

Знання і навички, отримані при вивченні дисципліни, будуть використовуватись у професійному контексті фахівцями з розробки математичного, інформаційного та програмного забезпечення інформаційних систем, у галузі інформаційних технологій, а також адміністраторами баз даних і систем, а також, при виконанні наукових досліджень у разі навчання за галузях 12 «Інформаційні технології», 15 «Автоматизація та приладобудування», 11 «Математика та статистика».

- Результати навчання:**
- РН03. Формулювати і перевіряти гіпотези; використовувати для обґрунтування висновків належні докази, зокрема, результати теоретичного аналізу, експериментальних досліджень і математичного та/або комп'ютерного моделювання, наявні літературні дані.
- РН04. Розробляти та досліджувати концептуальні, математичні і комп'ютерні моделі процесів і систем, ефективно використовувати їх для отримання нових знань та/або створення інноваційних продуктів у комп'ютерних науках та дотичних міждисциплінарних напрямках.
- РН05. Планувати і виконувати експериментальні та/або теоретичні дослідження з комп'ютерних наук та дотичних міждисциплінарних напрямів з використанням сучасних інструментів, критично аналізувати результати власних досліджень і результати інших дослідників у контексті усього комплексу сучасних знань щодо досліджуваної проблеми.
- РН06. Застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи.
- РН08. Визначати актуальні наукові та практичні проблеми у сфері комп'ютерних наук, глибоко розуміти загальні принципи та методи комп'ютерних наук, а також методологію наукових досліджень, застосувати їх у власних дослідженнях у сфері комп'ютерних наук та у викладацькій практиці.
- РН010. Відшуковувати, оцінювати та критично аналізувати інформацію щодо поточного стану та трендів розвитку, інструментів та методів досліджень, наукових та інноваційних проєктів з комп'ютерних наук.

Передумови до початку навчання:

Базові знання та уявлення з вищої математики, а саме розділи з числових методів вирішення рівнянь та інтеграли; знання дискретної математики, а саме розділи з теорії множин, їх відношення та відображення; знання з розробки та аналізу комп'ютерних алгоритмів, а саме розділи з побудови та аналізу алгоритмів, структур даних; знання теорії ймовірності, ймовірнісних процесів та математичної статистики, а саме розділи з перевірки статистичних гіпотез і статистичного оцінювання параметрів; з програмування, а саме розділи з об'єктно-орієнтованої методології побудови програм, основ візуального програмування в середовищі Visual Studio 2010, основ програмування на мові Python.

Вивчення даного курсу також базується на матеріалах дисциплін "Методи інтелектуального аналізу даних", "Теорія та інженерія комплексних систем" підготовки здобувачів ступеня вищої освіти «Доктор філософії» за спеціальністю 122 «Комп'ютерні науки».

Мета курсу (набуті компетентності)

В наслідок вивчення даного навчального курсу здобувач вищої освіти набуде наступних компетентностей:

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

СК01. Здатність виконувати оригінальні дослідження, досягати наукових результатів, які створюють нові знання у комп'ютерних науках та дотичних до них міждисциплінарних напрямках і можуть бути опубліковані у провідних наукових виданнях з комп'ютерних наук та суміжних галузей.

СК03. Здатність виявляти, ставити та вирішувати дослідницькі науково-прикладні задачі та/або проблеми в сфері комп'ютерних наук, оцінювати та забезпечувати якість виконуваних досліджень.

СК04. Здатність ініціювати, розробляти і реалізовувати комплексні інноваційні проекти у галузі комп'ютерних наук та дотичні до неї міждисциплінарних проектах, демонструвати лідерство під час їх реалізації.

СК08. Здатність використовувати можливості систем штучного інтелекту для аналізу наукових проблем та розробки адаптивних систем керування.

Структура курсу

№	Тема	Години (ЛК/ЛБ/ПЗ)	Стислий зміст	Інструменти і завдання
1.	Вступ у штучний інтелект та машинне навчання	2/0/2	Що таке штучний інтелект (ШІ) та машинне навчання (МН). Історія та основні автори в галузі. Різниця між ШІ та МН.	Участь в обговоренні
2.	Галузі застосування машинного навчання	2/0/1	Сфери використання: бізнес, медицина, освіта, технології, фінанси. Розгляд практичних прикладів застосування МН у реальних проектах.	Участь в Обговоренні Індивідуальні завдання
3.	Джерела та типи даних для машинного навчання	2/0/1	Джерела даних та різновиди даних: структуровані, неструктуровані, часові ряди.	Участь в обговоренні
4.	Методи машинного навчання	2/0/1	Контрольоване та неконтрольоване навчання. Основні методи: класифікація, регресія, кластеризація, асоціативний аналіз.	Участь в обговоренні Індивідуальні завдання
5.	Програмне забезпечення для машинного навчання	2/0/1	Огляд основних платформ та бібліотек: TensorFlow, PyTorch, scikit-learn. Інструменти для візуалізації даних.	Участь в обговоренні
6.	Алгоритми машинного навчання	2/0/1	Топ-10 алгоритмів: лінійна регресія, класифікація, метод опорних векторів, дерева рішень, кластеризація. Особливості та вибір алгоритмів для різних задач.	Участь в обговоренні
7.	Процес машинного навчання	2/0/1	Технологія CRISP-DM: процес побудови моделі від збору даних до оцінки результатів. Підготовка даних, моделювання, оцінка та вдосконалення моделей.	Участь в Обговоренні Індивідуальні завдання
8.	Джерела помилок у машинному навчанні	2/0/1	Зсув (Bias) та розкид (Variance): концепції та причини виникнення помилок. Перенавчання та недонавчання.	Участь в обговоренні
9.	Контрольоване навчання	2/0/1	Класифікація та регресія. Основні алгоритми: метод опорних векторів (SVM), дерева рішень, лінійна регресія.	Участь в обговоренні
10.	Неконтрольоване навчання	2/0/2	Кластеризація, виявлення аномалій, зниження розмірності. Використання неконтрольованих методів для аналізу даних.	Участь в Обговоренні Тести

Рекомендована література

1. I. Skarga-Bandurova, T. Biloborodova Foundations of Data Science for IoT and IoE Internet of Things for Industry and Human Applications. Volume 1. Fundamentals and Technologies. Data Science for IoT and IoE. Section 5. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, 608 p., pp. 206-235
2. I. Skarga-Bandurova, T. Biloborodova Data Mining and Processing for IoT and IoE Internet of Things for Industry and Human Applications. Volume 1. Fundamentals and Technologies. Data Science for IoT and IoE. Section 6. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, 608 p., pp. 236-267
3. Skarga-Bandurova I.S., Gorbenko A.V., Biloborodova T.O., Koval V.S., Sachenko A.O., Tarasyuk O. M. Data Science for Internet of Things and Internet of Everything: Practicum Skarga-Bandurova I.S. and Gorbenko A.V. (Eds.) – Ministry of Education and Science of Ukraine, Volodymyr Dahl East Ukrainian National University, National Aerospace University “Kharkiv Aviation Institute”, Ternopil National Economic University, 2019. – 169 p.
4. I. Skarga-Bandurova, T. Biloborodova Wearable and Embedded IoT-based Solutions for Biomedical Applications Internet of Things for Industry and Human Applications. Volume 3. Assessment and Implementation. IoT for Healthcare systems. Section 46. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, pp. 535-575.
5. I. Skarga-Bandurova, T. Biloborodova Devices with Reconfigurable Architecture for Biomedical IoT-based Applications Internet of Things for Industry and Human Applications. Volume 3. Assessment and Implementation. IoT for Healthcare systems. Section 47. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, pp. 576-596.
6. I. Skarga-Bandurova, Y. Krytska IoT based Water Quality Monitoring System in Internet of Things for Industry and Human Applications. Volume 3. Assessment and Implementation. IoT for Ecology, Safety and Security Monitoring Systems. Section 49. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, pp. 627-671
7. I. Skarga-Bandurova, A. Velykzhanin Basic concepts and approaches to development and implementation of IoT in Internet of Things for Industry and Human Applications. Volume 2. Modelling and Development. Development and implementation of IoT based systems. Section 28. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, pp. 403-435

8. I. Skarga-Bandurova, M. Nesterov, A. Velykzhanin, Algorithms and Applications for Utilization of SDN Technology to IoT in Internet of Things for Industry and Human Applications. Volume 2. Modelling and Development. Software Defined Networks and IoT. Section 22./ Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, pp. 194-240.
9. I. Skarga-Bandurova, M. Derkach IoT for Public Transport Information Service Delivering in Internet of Things for Industry and Human Applications. Volume 3. Assessment and Implementation. IoT for Intelligent Transportation Systems. Section 41. Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, pp. 371-401.
10. Soldatos, John, ed. Building Blocks for IoT Analytics. River Publishers, 2016.
11. Soni, N., Sharma, E. K., Singh, N., & Kapoor, A. (2020). Artificial Intelligence in Business: From Research and Innovation to Market Deployment. Procedia Computer Science, 167, 2200-2210.

Методичне забезпечення

1. I. Skarga-Bandurova, T. Biloborodova Foundations of Data Science for IoT and IoE Internet of Things for Industry and Human Applications. Volume 1. Fundamentals and Technologies. Data Science for IoT and IoE. Section 5. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, 608 p., pp. 206-235
2. I. Skarga-Bandurova, T. Biloborodova Data Mining and Processing for IoT and IoE Internet of Things for Industry and Human Applications. Volume 1. Fundamentals and Technologies. Data Science for IoT and IoE. Section 6. / Ed. V. S. Kharchenko. – Ministry of Education and Science of Ukraine, National Aerospace University KhAI, 2019, 608 p., pp. 236-267
3. Skarga-Bandurova I.S., Gorbenko A.V., Biloborodova T.O., Koval V.S., Sachenko A.O., Tarasyuk O. M. Data Science for Internet of Things and Internet of Everything: Practicum Skarga-Bandurova I.S. and Gorbenko A.V. (Eds.) – Ministry of Education and Science of Ukraine, Volodymyr Dahl East Ukrainian National University, National Aerospace University “Kharkiv Aviation Institute”, Ternopil National Economic University, 2019. – 169 p.

Стандарти за курсом

1. Національний інститут стандартів і технологій (NIST), Стандарти науки про дані та аналізу великих даних: <https://bigdatawg.nist.gov/standards/>

2. Група спеціальних інтересів з виявлення знань та інтелектуального аналізу даних (SIGKDD) Асоціації обчислювальної техніки (Association for Computing Machinery, ACM): <https://www.kdd.org/>
3. Процес міжгалузевого стандарту для інтелектуального аналізу даних (CRISP-DM): <https://www.datascience-pm.com/crisp-dm-2/>
4. Data Mining Group (DMG): <https://www.dmg.org/>
5. Стандарти мови розмітки предиктивних моделей (PMML): <http://dmg.org/pmml/v4-3/GeneralStructure.html>

Оцінювання курсу

За повністю виконані завдання студент може отримати визначену кількість балів:

Інструменти і завдання	Кількість балів
Участь в обговоренні	10
Тести	15
Індивідуальні завдання	45
Іспит	30
Разом	100

Шкала оцінювання студентів

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
82-89	B	добре	
74-81	C		
64-73	D	задовільно	
60-63	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

Політика курсу

Плагіат та академічна доброчесність:

Під час виконання завдань здобувач вищої освіти має дотримуватись політики академічної доброчесності. Запозичення не мають перевищувати 20% від загального обсягу роботи і мають бути оформлені відповідними посиланнями.

Завдання і заняття:

Аудиторні заняття мають відвідуватись регулярно. Пропущені заняття (з будь-яких причин) мають бути відпрацьовані з отриманням відповідної оцінки не пізніше останнього тижня поточного семестру. В разі поважної причини (хвороба, академічна мобільність тощо) терміни можуть бути збільшені за письмовим дозволом декана.

Здобувач вищої освіти може пройти певні онлайн-курси, які пов'язані з темами дисципліни, на онлайн-платформах. При поданні документу про проходження курсу студенту можуть бути перезараховані певні теми курсу та нараховані бали за завдання.

Всі завдання, передбачені програмою курсу мають бути виконані своєчасно і оцінені в спосіб, зазначений вище

Поведінка в аудиторії:

На заняття здобувачі вищої освіти вчасно приходять до аудиторії відповідно до діючого розкладу та обов'язково мають дотримуватися вимог техніки безпеки.

Під час занять здобувачі вищої освіти:

- не залишають аудиторію без дозволу викладача;
- не заважають викладачу проводити заняття.

Під час контролю знань здобувачі вищої освіти:

- є підготовленими відповідно до вимог даного курсу;
- розраховують тільки на власні знання;
- виконують усі вимоги викладачів щодо контролю знань.