



Силабус освітнього компонента

Програма навчальної дисципліни



Інженерія машинного навчання

Шифр та назва спеціальності
113 – Прикладна математика

Інститут
ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма
Інтелектуальний аналіз даних

Кафедра
Комп'ютерна математика і аналіз даних

Рівень освіти
Магістр

Тип дисципліни
Профільна, Вибіркова

Семестр
2

Мова викладання
Українська

Викладачі, розробники



Колбасін Владислав Олександрович

vladyslav.kolbasin@khai.edu.ua

Старший викладач кафедри комп'ютерної математики і аналізу даних.

Досвід роботи – 15 років. Провідний лектор з дисциплін: «Об'єктно орієнтоване програмування», «Комп'ютерна геометрія і графіка».

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна спрямована на оволодіння теоретичними та практичними основами інженерії машинного навчання. Розглянуто задачі, які виникають у реальних проектах, що пов'язані з використанням машинного навчання.

Мета та цілі дисципліни

Метою дисципліни є формування уявлень про основні етапи, задачі, та методи вирішення проблем, що виникають у проектах, що пов'язані з використанням машинного навчання.

Формат занять

Лекції, лабораторні заняття, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

ЗК 3. Здатність до безперервного навчання, придбання нових знань і умінь, у тому числі в галузі, відмінній від професійної.

ЗК 4. Здатність виявляти, ставити та вирішувати проблеми у професійній діяльності.

ЗК 5. Здатність генерувати нові ідеї (креативність) і нестандартні підходи до їхньої реалізації, гнучке адаптування до реальних професійних ситуацій, проявляти творчий підхід, ініціативу.

ЗК 6. Здатність критично оцінювати й переосмислювати накопичений досвід (власний і чужий),

аналізувати свою професійну і соціальну діяльність.

ЗК 7. Здатність працювати з інформацією: знаходити і використовувати інформацію з різних джерел, потрібну для розв'язання професійних завдань.

СК 1. Здатність формулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, перевіряти коректність постановки, у тому числі в умовах невизначеності.

СК 2. Здатність обирати, розробляти та досліджувати математичний аналітичний або чисельний метод розв'язання практичних задач, що забезпечує потрібні точність і надійність результату.

СК 4. Здатність розробляти алгоритми аналізу невизначених великих даних, розробляти відповідні програмні засоби та документацію, проектувати програмні системи, бази даних і знань.

СК 5. Здатність до проведення математичного і комп'ютерного моделювання та обчислювального експерименту, збору, візуалізації, аналізу та обробки отриманих даних, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів.

СК 11. Здатність розробляти, досліджувати та застосовувати математичні методи й алгоритми машинного навчання, м'яких обчислень і обчислювального інтелекту для аналізу невизначених даних, прогнозування та прийняття рішень.

СК 12. Здатність до розробки й експлуатації спеціалізованих програмних засобів інтелектуального аналізу даних, текстів, сигналів і зображень.

СК 13. Здатність до розробки та експлуатації спеціалізованих програмних засобів обробки великих масивів даних на основі інформаційних технологій розподілених і хмарних обчислень.

Результати навчання

РН 1. Демонструвати знання і розуміння основних концепцій, принципів, теорій фундаментальної та прикладної математики і використовувати їх на практиці.

РН 2. Уміти формалізувати задачі, сформульовані мовою певної предметної галузі й обирати раціональний метод вирішення; розв'язувати задачі аналітичними або чисельними методами, оцінювати точність і достовірність отриманих результатів та виконувати їхню інтерпретацію.

РН 3. Володіти методами розробки, дослідження та застосування математичних моделей складних об'єктів і процесів, у тому числі із застосуванням методів обчислювального інтелекту.

РН 7. Уміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символічних алгоритмів.

РН 12. Знати і розуміти сучасні методи розв'язання математичних задач статистичного й інтелектуального аналізу даних, прогнозування тощо.

РН 13. Знати і розуміти методи розв'язання математичних задач інтелектуального інформаційного пошуку та видобування знань.

РН 14. Уміти застосовувати наявні існуючі і розробляти нові алгоритми та програмні засоби для статистичного й інтелектуального аналізу невизначених даних.

РН 15. Уміти застосовувати наявні існуючі і розробляти нові алгоритми та програмні засоби обробки даних, текстів, сигналів і зображень.

РН 16. Уміти застосовувати сучасні інформаційні технології та програмне забезпечення для обробки великих масивів даних на основі розподілених і хмарних сервісів.

Обсяг дисципліни

Загальний обсяг дисципліни 150 год. (5 кредитів ECTS): лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 86 год.

Передумови вивчення дисципліни (пререквізити)

Для успішного проходження курсу необхідно мати знання та практичні навички з дисципліни «Математичні методи машинного навчання 1», а також з дисциплін профільного пакету ВП1.*, а саме, ВП1.1 – «Методи та технології роботи з великими даними», або ВП1.2 – «Аналіз графових даних».

Особливості дисципліни, методи та технології навчання

Потрібні навички програмування. Навчальні матеріали доступні студентам на сайті викладача.

Програма навчальної дисципліни

Теми лекційних занять

- Тема 1. Вступ до предмету. Загальна схема розробки проекту машинного навчання.
- Тема 2. Основи інженерії машинного навчання.
- Тема 3. Особливості тренувальних даних.
- Тема 4. Інженерія признаков (feature engineering).
- Тема 5. Утиліта Data Version Control.
- Тема 6. Утиліта MLFlow.
- Тема 7. Утиліта Weights & Bias.
- Тема 8. Вибір моделей машинного навчання, побудова ансамблей моделей.
- Тема 9. Витік даних (data leakage). Методи оцінки важливості признаков моделей.
- Тема 10. Методи візуалізації та інтерпретації роботи нейронних мереж.
- Тема 11. Методи оцінки якості моделей машинного навчання.
- Тема 12. Квантизація моделей.
- Тема 13. Методи розгортання застосунків (deployment).
- Тема 14. Відтворюваність експериментів.
- Тема 15. Змагальні атаки (adversarial attacks) на моделі машинного навчання.
- Тема 16. Огляд курсу та обговорення досягнутих результатів.

Теми лабораторних робіт

- Тема 1. Розробка простого пайплайну машинного навчання.
- Тема 2. Сегментація пайплайну машинного навчання на кроки.
- Тема 3. Розробка DVC pipeline
- Тема 4. Використання MLFlow для трекінгу тренування моделей.
- Тема 5. Використання W&B для трекінгу тренування моделей.

Самостійна робота

Під час самостійної роботи студенти вивчають лекційний матеріал, виконують лабораторні завдання, готуються до контрольних робіт, іспиту.

Література та навчальні матеріали

Основна література

1. Chip Huyen. Designing Machine Learning Systems. O'Reilly Media, Inc. – 2022. – ISBN: 9781098107963.
2. Andrew P. McMahon, Adi Polak . Machine Learning Engineering with Python: Manage the lifecycle of machine learning models using MLOps with practical examples. Packt Publishing; 2nd edition – 2023. – ISBN 978-1837631964.
3. <https://www.datascience-pm.com/crisp-dm-2/>
4. MLOps: Continuous delivery and automation pipelines in machine learning.
<https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning>
5. <https://cs329s.stanford.edu>
6. <https://dvc.org/doc>
7. <https://mlflow.org/>
8. <https://docs.wandb.ai/guide>
9. Attacking Machine Learning with Adversarial Examples.
<https://openai.com/blog/adversarial-example-research/>

Додаткова література

10. How ML Breaks: A Decade of Outages for One Large ML Pipeline.
<https://www.youtube.com/watch?v=hBMHohkRgAA>
11. <https://realpython.com/python-data-version-control/>
12. <http://mlwave.com/kaggle-ensembling-guide/>
13. <https://scikit-learn.org/stable/modules/ensemble.html>

14. <https://www.youtube.com/watch?v=dWhdWxgt5SU>
15. "Interpretable Machine Learning" book.
<https://christophm.github.io/interpretable-ml-book/>
16. A Survey Of Methods For Explaining Black Box Models.
<https://arxiv.org/pdf/1802.01933.pdf>
17. Techniques for Interpretable Machine Learning.
<https://arxiv.org/pdf/1808.00033.pdf>
18. <https://www.kaggle.com/code/samueltortinhas/weights-biases-tutorial-beginner>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Необхідною умовою здачі заліку або іспиту є виконання лабораторних робіт.
За написання контрольних робіт нараховується 30 балів.
Здача лабораторних – 30 балів.
Іспит – 40 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 «ХП»: 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 «ХП» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>.

Погодження

Силабус погоджено

Дата погодження, підпис
31.08.2023 р.



Завідувач кафедри
Олена АХІЗЕР

Дата погодження, підпис
31.08.2023 р.



Гарант ОП
Олексій ГАЛУЗА