

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «МАШИННЕ НАВЧАННЯ»



| | |
|-----------------------|--|
| Ступінь освіти | магістр |
| Спеціальність | 121 Інженерія програмного забезпечення |
| Освітня програма | Інженерія програмного забезпечення |
| Тривалість викладання | осінній семестр (1, 2 чверті) |
| Кількість кредитів | 5 кредитів ЄКТС (150 годин) |
| Заняття: | |
| лекції: | 2 год/тиждень |
| лабораторні: | 2 години/тиждень |
| Мова викладання | українська |

Сторінка курсу в СДО НТУ «ДП»: <https://do.nmu.org.ua/course/view.php?id=4872>

Кафедра, що викладає Програмного забезпечення комп'ютерних систем



Викладач:

Приходченко Сергій Дмитрович
Доцент, к.т.н

Персональна сторінка

<https://pzks.nmu.org.ua/ua/teachers/teachers.php>
<https://scholar.google.com/citations?user=Aqq63hMAAAAJ>
E-mail: Prykhodchenko.s.d@nmu.one

1. Анотація до курсу

Дисципліна спрямована на формування теоретичних знань і практичних навичок з курсу машинного навчання, включаючи контрольоване навчання (множинна лінійна регресія, логістична регресія, нейронні мережі та дерева рішень), неконтрольоване навчання (кластеризація, зменшення розмірності, системи рекомендацій). У курсі представлені практики, які використовуються в Silicon Valley для інновацій у сфері штучного інтелекту та машинного навчання (оцінка та налаштування моделей, орієнтований на дані підхід до покращення продуктивності тощо). У рамках дисципліни студенти опановують ключові концепції та практичні навички швидкого та потужного застосування машинного навчання для вирішення складних проблем реального світу.

2. Мета та завдання курсу

Мета дисципліни – вивчення головних аспектів відгалуження напрямку штучного інтелекту, а саме машинного навчання, що спеціалізується на навчанні математичних моделей, аналізу їх функціонування, та моделювання структур, що є найбільш придатними для аналізу тих або інших прикладних завдань штучного інтелекту.

Завдання курсу:

Навчити здобувачів вищої освіти:

- обґрунтовувати роль і значення Data Science та Machine Learning щодо сутності об'єктів і процесів даних, будови програмних систем з їх застосуванням;
- визначати основні сценарії використання і застосування Data Science та Machine Learning, характеристики алгоритмів Machine Learning;
- розрізняти парадигми застосування регресії і класифікації та оцінювати точності навчених моделей;
- аналізувати первинну інформацію, що отримана з різних джерел, в тому числі набори даних, визначати найбільш придатний алгоритм;
- користуватися отриманою інформацією в своїй майбутній професії.

3. Результати навчання

Дисциплінарні результати навчання:

- досліджувати моделі інформаційних процесів в області машинного навчання;
- класифікувати данні для машинного навчання;
- застосовувати на практиці сучасні засоби розроблення програмного забезпечення з елементами та алгоритмами машинного навчання;
- розрізняти парадигми застосування регресії і класифікації та оцінювати точності навчених моделей;
- модифікувати існуючі та розробляти нові сценарії використання і застосування машинного навчання;
- прогнозувати рішення машинного навчання.

Дисциплінарні результати навчання сформовано на основі ПРН освітньо-професійної програми «Інженерія програмного забезпечення» другого (магістерського) рівня вищої освіти (РН03, РН04, РН09, РН10, РН14).

4. Структура курсу

| Види та тематика навчальних занять | Внесок в загальну оцінку, % |
|---|-----------------------------|
| ЛЕКЦІЇ | 42 |
| Тема 1. Основи мови програмування Python Поняття алгоритму. Особливості застосування мови програмування Python. Алгоритмічні структури в мові Python. Модульність у програмі: функції та модулі. Типи даних Python. Сучасні приклади застосування. | |

| Види та тематика навчальних занять | Внесок в загальну оцінку, % |
|---|-----------------------------|
| Перспективи розвитку. Розвиток галузі. | |
| Тема 2. Введення в Data Science та Machine Learning Вступ в Data Science і Machine Learning. Базові поняття Big Data, Business Intelligence, Data Mining, Data Science, Machine Learning, Artificial Intelligence. Історія розвитку Data Science, сучасні перспективи. Особливості застосування Data Science. Огляд процесу Data Science проекту. Сценарії використання і застосування ML в сучасному світі. | |
| Тема 3. Регресія і класифікація. Лінійна регресія. Поліноміальна регресія Методи побудови дерев рішень. Метод найближчих сусідів. Метрики оцінки якості регресії. Покращення якості регресійних моделей. Алгоритм покриття. Алгоритм CART. Байєсовські методи класифікації. Метод опорних векторів: лінійний і нелінійний випадки. | |
| <i>Тестова контрольна робота №1 (за темами 1-3)</i> | 21 |
| Тема 4. Кластеризація та рекомендаційні алгоритми Основні алгоритми кластеризації. Особливості застосування алгоритмів кластеризації. Метод k-середніх. Ієрархічний кластерний аналіз. Кореспонденс-аналіз. | |
| Тема 5. Оцінка точності навчених моделей, вибір кращої Методи поліпшення якості моделей машинного навчання. Аугментація. Feature engineering. Cross-validation. Боротьба з перенавчанням. Боротьба з дисбалансом даних. Регуляризація моделі. | |
| Тема 6. Часові ряди і прогнозування подій Теоретичні аспекти прогнозування часових рядів. Основні методи та моделі прогнозування часових рядів. AR, MA, ARMA. ARIMA. VAR. Lstm. | |
| <i>Тестова контрольна робота №2 (за темами 4-6)</i> | 21 |
| ЛАБОРАТОРНІ ЗАНЯТТЯ | 58 |
| Лабораторна робота 1. Основи мови програмування Python. | |

| Види та тематика навчальних занять | Внесок в загальну оцінку, % |
|---|------------------------------------|
| Підключення стандартних бібліотек. Програмування пробного додатку. | |
| <i>Звіт з роботи №1 та захист лабораторної роботи.</i> | 8 |
| Лабораторна робота 2. Введення в Data Science та Machine Learning. Ініціалізація. Основні елементи. | |
| <i>Звіт з роботи №2 та захист лабораторної роботи.</i> | 10 |
| Лабораторна робота 3. Регресія і класифікація. Розбір методів регресії та класифікації. Програмування додатку. | |
| <i>Звіт з роботи №3 та захист лабораторної роботи.</i> | 10 |
| Лабораторна робота 4. Кластеризація та рекомендаційні алгоритми. Базові налаштування. | |
| <i>Звіт з роботи №4 та захист лабораторної роботи</i> | 10 |
| Лабораторна робота 5. Оцінка точності навчених моделей, вибір кращої. | |
| <i>Звіт з роботи №5 та захист лабораторної роботи.</i> | 10 |
| Лабораторна робота 6. Часові ряди і прогнозування подій. | |
| <i>Звіт з роботи №6 та захист лабораторної роботи.</i> | 10 |
| Загальна кількість балів | 100 |

5. 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, мова програмування Python.

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 |

Загальні критерії досягнення результатів навчання відповідають описам 7-го кваліфікаційного рівня НРК.

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 – 42 бали та вираховується відсоток кожної (розподіл % за окремими контрольними роботами див. в таблиці розділу 4). Загалом за дві контрольні тестові роботи отримується **максимум 42 бали**, тобто 42% від оцінки за дисципліну.

Лабораторні роботи (шість робіт – у вигляді індивідуального завдання з кожної, 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 58% (максимум 58 балів). При несвоєчасному здаванні роботи оцінка знижується вдвічі. Лабораторні роботи захищаються у вигляді опитування за звітом, і захист враховується, як 50% від оцінки за роботу. У сумі за лабораторну частину курсу при поточному оцінюванні отримується **максимум 58 балів**.

Отримані бали за теоретичну частину та лабораторні роботи додаються і є підсумковою оцінкою 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 |
|--------------------|-------------------|------------|
| 42 | 58 | 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 (іспит)** під час сесії.

Іспит проводиться у вигляді комплексної контрольної роботи, яка включає запитання з теоретичної та практичної частини курсу. Білет складається з **30 тестових завдань** з чотирма варіантами відповідей, одна правильна відповідь оцінюється в 2 бали (**разом 60 балів**) та **2 завдань** з практичної частини, кожне з запитань оцінюється максимум у 20 балів (**разом 40 балів**), причому:

- 20 балів – відповідність еталону;
- 15 балів – відповідність еталону з незначними помилками;
- 10 балів – часткова відповідність еталону, питання повністю не розкриті;
- 5 балів – 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 (вигадування 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://bit.ly/3ExtVKY>).

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.

7.5. Відвідування занять. Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, академічна мобільність, які необхідно підтверджувати документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту.

За об'єктивних причин (наприклад, міжнародна мобільність) навчання може відбуватись в он-лайн формі за погодженням з керівником курсу.

7.6. Опитування. Наприкінці вивчення курсу та перед початком сесії здобувача вищої освіти буде запропоновано анонімно заповнити електронні анкети (Microsoft Forms Office 365), які буде розіслано на ваші університетські поштові скриньки. Заповнення анкет є важливою складовою вашої навчальної активності, що дозволить оцінити дієвість застосованих методів викладання та врахувати ваші пропозиції стосовно покращення змісту навчальної дисципліни.

8 Рекомендовані джерела інформації

Базові

1. Кононова К.Ю. Машинне навчання: методи та моделі / Підручник/ – Х., ХНУ імені В.Н. Каразіна, 300 с., 2020р.
2. Матеріали Массачусетського технологічного університету з дисципліни «Introduction to Machine learning» (<https://cutt.ly/VWHEFCL>).
3. Tom M. Mitchell. Machine Learning [<http://www.cs.cmu.edu/~tom/mlbook.html>].
4. Stephen Marsland. Machine Learning: An Algorithmic Perspective, 452 p., 2015.
5. Wilson, Halsey (2018). Artificial intelligence. Grey House Publishing. ISBN 978-1682178676.

6. Charu C. Aggarwal. Neural Networks and Deep Learning. - Springer International Publishing, 2018. - XXIII, 497pp/ 128 b/w illustrations, 11 illustrations in colour.
7. Hitoshi Iba, Nasimul Noman. Deep Neural Evolution. Deep Learning with Evolutionary Computation. - Springer Singapore, 2020. - 438 pp/ 114 b/w illustrations, 107 illustrations in colour.
8. Sandro Skansi. Introduction to Deep Learning. From Logical Calculus to Artificial Intelligence/. - Springer International Publishing, 2018. – 191 pp/ 38 b/w illustrations.
9. Hamed Habibi Aghdam, Elnaz Jahani Heravi. Guide to Convolutional Neural Networks. A Practical Application to Traffic-Sign Detection and Classification. — Springer International Publishing, 2017. — 282 pp./ 39 b/w illustrations, 111 illustrations in colour.
10. Najafabadi, M., Villanustre, F., Khoshgoftaar, T., Seliya, N., Wald, R. and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1).

Додаткові

11. Kuen, J., Lim, K. and Lee, C. (2015). Self-taught learning of a deep invariant representation for visual tracking via temporal slowness principle. Pattern Recognition.
12. M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” 2016.
13. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT Press: Cambridge, MA, USA, 2016.
14. Al-Rfou, R.; Alain, G.; Almahairi, A.; Angermueller, C.; Bahdanau, D.; Ballas, N.; Bastien, F.; Bayer, J.; Belikov, A.; Belopolsky, A.; et al. Theano: A Python framework for fast computation of mathematical expressions. arXiv 2016, arXiv:abs/1605.02688.
15. The latest in machine learning. Papers With Code [Електронний ресурс]. – Електрон. дан. – Режим доступу: World Wide Web. – URL: <https://paperswithcode.com/>.
16. THE NEURAL NETWORK ZOO. [Електронний ресурс]. – Електрон. дан. – Режим доступу: World Wide Web. – URL: <https://www.asimovinstitute.org/neural-network-zoo/>.