

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВОДНОГО ГОСПОДАРСТВА ТА
ПРИРОДОКОРИСТУВАННЯ**

Навчально-науковий інститут кібернетики, інформаційних технологій та інженерії

04-05-241S

СИЛАБУС	Методи машинного навчання	
SYLLABUS	Machine Learning Techniques	
Шифр за ОП	ПП 6	
Code in Degree Programme		
Освітній рівень	магістерський (другий)	
Level of Education	Master's (second)	
Галузь знань	12	Інформаційні технології
Field of Knowledge		Information technology
Спеціальність	126	Інформаційні системи та технології
Field of Study		Information systems and technologies
Освітня програма	Інформаційні технології в бізнесі	
Degree Programme	Information technologies in business	

РІВНЕ – 2024

Силабус навчальної дисципліни «Методи машинного навчання» для здобувачів вищої освіти ступеня «магістр», які навчаються за освітньо-професійною програмою «Інформаційні технології в бізнесі» спеціальності 126 «Інформаційні системи та технології». Рівне. НУВГП. 2024. 12 стор.

ОП на сайті університету: <https://ep3.nuwm.edu.ua/30341/>

Розробник силабусу: *Грицюк П. М., д.е.н., професор, завідувач кафедри комп'ютерних технологій та економічної кібернетики*

Силабус схвалений на засіданні кафедри комп'ютерних технологій та економічної кібернетики

Протокол № 1 від “27” серпня 2024 року

Завідувач кафедри: *Грицюк П. М., д.е.н., професор.*

Керівник (гарант) ОП: *Барановський С. В., к.т.н, доцент, доцент кафедри комп'ютерних технологій та економічної кібернетики*

Схвалено науково-методичною радою з якості ННІ КІТІ


Протокол № 9 від “30” серпня 2024 року

Голова науково-методичної ради з якості ННІ КІТІ:
Мартинюк П. М., д.т.н., професор

Попередня версія силабусу (вказати шифр) 04-05-232S

© НУВГП, 2024

ПРОГРАМА навчальної дисципліни «Методи машинного навчання»

<p>Лектор</p> 	<p>Грицюк Петро Михайлович, доктор економічних наук, кандидат фізико-математичних наук, професор, завідувач кафедри комп'ютерних технологій та економічної кібернетики</p>
<p>Вікіситет</p>	<p>http://wiki.nuwm.edu.ua/index.php/Грицюк_Петро_Михайлович</p>
<p>ORCID</p>	<p>https://orcid.org/0000-0002-3683-4766</p>
<p>Як комунікувати</p>	<p>p.m.hrytsiuk@nuwm.edu.ua</p>
<p>ІНФОРМАЦІЯ ПРО ОСВІТНІЙ КОМПОНЕНТ</p>	
<p>Мета та завдання</p>	
<p>Метою вивчення дисципліни є формування у майбутніх фахівців знань та вмінь застосування сучасних методів та засобів розробки, дослідження та використання сучасних технологій обробки даних для вирішення задач класифікації, регресійного аналізу, прогнозування та ухвалення рішень.</p> <p>Основні завдання. Засвоєння теоретичних знань:</p> <ul style="list-style-type: none"> - теоретичних основ базових технологій машинного навчання; - методів, алгоритмів роботи, засобів реалізації та технологій налагодження систем обробки даних з використанням технологій обчислювального інтелекту; - сучасних програмних засобів реалізації технологій машинного навчання. <p>Набуття умінь:</p> <ul style="list-style-type: none"> - обґрунтовано вибирати конкретні технології та алгоритми машинного навчання при розв'язанні відповідних практичних задач; - здійснювати підготовку та первинну обробку даних для побудови моделей систем методами машинного навчання; - вирішувати задачі автоматизації підтримки рішень, розпізнавання образів, діагностики, класифікації, оптимізації та аналізу даних методами машинного навчання; - засвоїти технологію реалізації методів машинного навчання в середовищі Google Colab/Jupyter Notebook із застосуванням бібліотек Python. 	
<p>Посилання на розміщення освітнього компонента на навчальній платформі Moodle, на платформі освітніх програм та їхніх освітніх компонентів</p>	
<p>https://exam.nuwm.edu.ua/course/view.php?id=5947</p>	
<p>Передумови вивчення дисципліни (місце освітнього компонента в структурно-логічній схемі)</p>	

Основою для успішного засвоєння матеріалу дисципліни є знання основ дисциплін "Вища математика", "Теорія ймовірності і математична статистика", „Програмування", "Моделювання та аналіз бізнес-процесів". Отримані знання та вміння в результаті вивчення навчальної дисципліни «Методи машинного навчання» використовуються для подальшого освоєння такої освітньої компоненти як «Кваліфікаційна робота магістра».

Компетентності

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.
ЗК4. Здатність розробляти проекти та управляти ними.
СК1. Здатність розробляти та застосувати ІСТ, необхідні для розв'язання стратегічних і поточних задач.
СК4. Здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.
СК5. Здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.
СК7. Розробляти і реалізовувати інноваційні проекти у сфері ІСТ.

Програмні результати навчання (ПРН). Результати навчання (РН)

РН1. Відшукувати необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, аналізувати та оцінювати цю інформацію.
РН6. Обґрунтовувати вибір технічних та програмних рішень з урахуванням їх взаємодії та потенційного впливу на вирішення організаційних проблем, організувати їх впровадження та використання.
РН8. Розробляти моделі інформаційних процесів та систем різного класу, використовувати методи моделювання, формалізації, алгоритмізації та реалізації моделей з використанням сучасних комп'ютерних засобів.
РН11. Розв'язувати задачі цифрової трансформації у нових або невідомих середовищах на основі спеціалізованих концептуальних знань, що включають сучасні наукові здобутки у сфері інформаційних технологій, досліджень та інтеграції знань з різних галузей.

Структура та зміст освітнього компонента

Змістовий модуль 1

Тема 1. Засоби Python для аналізу даних

Аналіз даних. Програмні пакети та середовища : Jupyter Notebook, Google Colab. Пакет NumPy. Обробка масивів та фреймів. Пакет Pandas. Візуалізація даних. Пакет Matplotlib. Пакет Scikit-learn. Кількість годин: лекції – 2, лабораторні – 4. Результати навчання: РН6, РН11.

Тема 2. Основні положення теорії машинного навчання

Мета машинного навчання. Задачі машинного навчання. Навчання з учителем і без учителя. Навчання з підкріпленням. Проблеми машинного навчання : недостатній обсяг даних; дані низької якості; несуттєві ознаки; перенавчання та недонавчання моделі. Кількість годин: лекції – 2 лабораторні – 0. Результати навчання: РН1.

Тема 3. Передобробка даних. Налаштування моделі

Завантаження даних в Google Colab. Типи даних. Візуалізація даних. Пошук зв'язків. Очищення даних. Масштабування даних. Оцінка моделі. Критерії якості моделі. Точне налаштування моделі. Решітковий пошук. Кількість годин: лекції – 2, лабораторні – 2. Результати навчання: РН1.

Тема 4. Методи відновлення регресії

Кореляція ознак. Задача лінійної регресії. Навчальна, контрольна і тестова вибірка. Навчання моделі. Оцінка якості моделі. Градієнтний спуск. Стохастичний градієнтний спуск. Поліноміальна регресія. Гребнева регресія. Лассо-регресія. Кількість годин: лекції – 4, лабораторні – 4. Результати навчання: РН8.

Тема 5. Методи класифікації. Логістична регресія

Задача класифікації. Бінарна та множинна класифікація. Бінаризація числових даних. Навчання двійкового класифікатора. Оцінка класифікатора. Матриця неточностей. Точність і чутливість моделі. Крива ROC. Використання перехресної перевірки. Класифікація з багатьма класами. Кількість годин: лекції – 3, лабораторні – 2. Результати навчання: РН8.

Змістовий модуль 2

Тема 6. Метод опорних векторів. Метод k-NN

Метод опорних векторів. Лінійна класифікація SVM. Нелінійна класифікація SVM. Гауссове ядро RBF. Регресія SVM. Функції рішень і прогнози. Метод k-NN. Кількість годин: лекції – 2, лабораторні – 4. Результати навчання: РН8.

Тема 7. Древа прийняття рішень

Древа прийняття рішень. Навчання та візуалізація ДПР. Оцінювання ймовірності класів. Алгоритм навчання CART. Обчислювальна складність. Обрізка дерева. Гіперпараметри регуляризації. Правила прийняття рішень. Регресія. Нестійкість. Кількість годин: лекції – 2, лабораторні – 2. Результати навчання: РН8.

Тема 8. Ансамблеве навчання

Ансамблеве навчання і випадковий ліс дерев. Класифікатори з вибором. Оцінка на невикористовуваних прикладах. Значущість ознак. Багінг і бустинг. Градієнтний бустинг. Стекінг. Кількість годин: лекції – 2, лабораторні – 2. Результати навчання: РН8, РН11.

Тема 9. Методи зниження розмірності

Навчання без учителя. Прокляття розмірності. Основні підходи до зниження розмірності. Алгоритм PCA. Інкрементний аналіз головних компонент. Рандомізований аналіз головних компонент. Ядерний аналіз головних компонент. Інші методи зниження розмірності. Кількість годин: лекції – 2, лабораторні – 2. Результати навчання: РН8, РН11.

Тема 10. Інші напрямки навчання без учителя

Кластеризація. Евристичні графові алгоритми. K-means. Ієрархічна кластеризація. Метод DBSCAN. Виявлення асоціацій. Правила асоціації. Підтримка, впевненість, часті набори товарів. Навчання з підкріпленням. Кількість годин: лекції – 3, лабораторні – 4. Результати навчання: РН11.

Індивідуальне навчально-дослідне завдання

Індивідуальне завдання визначає конкретну спрямованість роботи студента в сфері машинного навчання та уточнює склад і зміст інформаційного матеріалу, необхідного для опанування під час вивчення дисципліни.

Індивідуальне навчально-дослідне завдання з дисципліни «Методи машинного навчання» представлено у вигляді курсової роботи.

Тема індивідуального завдання визначається студентом, погоджується з викладачем, який може конкретизувати її, доповнити або змінити виходячи з потреб об'єкта курсової роботи.

ІНДЗ передбачає :

- описання вибраної області досліджень;
- пошук масиву реальних даних для вибраної області досліджень;
- попередню обробку та аналіз даних;
- побудову та навчання моделі регресії (класифікації, кластеризації);
- налаштування моделі;

- оцінку якості та точності побудованої моделі;
- висновки та рекомендації щодо застосування моделі для вирішення прикладних задач.

Виконання індивідуального завдання містить у собі виконання таких робіт, як вивчення вибраної області досліджень та пошук масиву реальних даних у цій області. Обсяг масиву повинен бути достатнім для використання методів машинного навчання. На наступному етапі слід сформулювати мету дослідження – який результат ми хочемо отримати по завершенню аналізу. Потім потрібно вибрати метод машинного навчання (регресія, класифікація, кластеризація), який найбільш підходить до вирішення поставленої задачі.

Початок реалізації проєкту передбачає попередню обробку даних. Цей етап передбачає перевірку пропущених даних, пошук викидів, стандартизацію/нормалізацію даних, візуалізацію даних. Потім потрібно провести попередній аналіз зв'язків між ознаками.

На наступному етапі потрібно побудувати та навчити модель машинного навчання, яка була вибрана для реалізації проєкту. При навчанні з учителем рекомендований обсяг навчальної вибірки – 70% від наявних даних. Потім слід оцінити якість/точність побудованої моделі та провести її оптимальне налаштування.

Завершальний етап передбачає висновки з проведеного дослідження та рекомендації щодо застосування моделі для вирішення прикладних задач.

Описання досліджуваної області, зібрані дані, результати їх обробки, алгоритм побудови і навчання моделі, оцінка якості моделі, висновки та рекомендації систематизуються по розділах і оформляються у вигляді пояснювальної записки до курсової роботи обсягом 20-25 сторінок. Список використаної літератури повинен містити не менше 10 джерел, включаючи англomовні статті, опубліковані не раніше 2015 року.

Форми та методи навчання

Методи та технології навчання: демонстрація, лекція, візуалізація, пошукова лабораторна робота, індивідуальне навчання, індивідуальне навчально-дослідне завдання, інтерактивні методи навчання.

Інструменти, обладнання, програмне забезпечення

Засоби навчання: технічні засоби, мультимедіа і проєкційна апаратура, комп'ютери, комп'ютерні системи та мережі, програмне забезпечення з відкритим доступом.

Порядок оцінювання програмних результатів навчання/результатів навчання

Підсумковий контроль відбувається у вигляді проходження двох модульних контролів у формі тестування на університетській платформі MOODLE. Дисципліна закінчується екзаменом, тому результати складання модульних контролів можуть зараховуватись як підсумковий контроль.

Модульний контроль проходить у формі тестування. Загальна кількість питань по кожному з модулів є не меншою від 300, з них у тест випадковим чином включаються 29 питань різного рівня складності з наступним розподілом балів:

- 1-й рівень: 22 питання по 0,5 бал кожне, разом 11 балів.
- 2-й рівень: 5 питань по 1.0 бал кожне, разом 5 балів.
- 3-й рівень: 2 питання по 2.0 бали кожне, разом 4 бали.

Контроль самостійної роботи проводиться на основі перевірки виконаних завдань. Усі форми контролю включено до 100-бальної шкали оцінки.

Поточний контроль (оцінювання результатів виконання лабораторних робіт, результатів самостійної роботи) здійснюється за такими вимогами:

- лабораторна робота повинна бути виконана згідно інструкції та з використанням програмного забезпечення, вказаного в інструкції;
- вхідні дані до лабораторної роботи повинні відповідати варіанту, який визначається номером студента у електронному журналі;
- звіт про лабораторну роботу повинен бути підготовленим з використанням текстового редактора Word;
- звіт про лабораторну роботу повинен містити тему роботи, змістовну постановку задачі, вхідні дані (або фрагмент вхідних даних), ключові фрагменти коду та основні результати, ілюстрації, висновки;
- студент повинен захистити лабораторну роботу на занятті чи консультації не пізніше 2 – 3 тижні після її виконання;
- захист лабораторної роботи відбувається у формі бесіди викладача та студента;
- захист відбувається в очній формі, або онлайн, якщо до цього змушують обставини;
- оцінка за лабораторну роботу виставляється за шкалою, наведеною в наступній таблиці і може бути знижена при наступних обставинах :
 - не всі завдання виконані правильно, вчасно і без зауважень;
 - студент має низький рівень знань по темі лабораторної роботи, не розуміє логіки програмного коду та технології його виконання;
 - захист відбувається із значним запізненням;
 - звіт про виконання лабораторної роботи не відповідає наведеному вище вимогам;
- якщо лабораторна робота виконана в основному (виконано $\geq 80\%$ завдань), зниження оцінки не може перевищувати 50%.

Шкала оцінювання навчальних досягнень здобувачів вищої освіти

Вид заняття	Бали
1. Поточна складова оцінювання	
1.1. Лабораторна робота 1. Середовище Google Colab. Передобробка даних. Пошук зв'язків	5
1.2. Лабораторна робота 2. Візуалізація даних з використанням пакетів Matplotlib та Seaborn	5
1.3. Лабораторна робота 3. Модель множинної лінійної регресії.	6
1.4. Лабораторна робота 4. Регуляризація моделі множинної лінійної регресії. Гребнева регресія. Лассо-регресія	5
1.5. Лабораторна робота 5. Перехресна крос-валідація моделі регресії.	5
1.6. Лабораторна робота 5. Бінарна класифікація зразків. Модель логістичної регресії	6

1.7. Лабораторна робота 6. Дерева прийняття рішень. Налаштування глибини дерева.	6
1.8. Лабораторна робота 7. Метод опорних векторів	6
1.9. Лабораторна робота 9. Методи кластеризації. Метод K-means. Ієрархічна модель. Модель гауссової суміші. Модель DBSCAN.	6
1.10. Лабораторна робота 10. Ансамблеве навчання.	6
1.11. Робота на лекціях	4
Всього поточна складова оцінювання:	60
2. Модульна складова оцінювання	
2.1. Модульний контроль №1	20
2.2. Модульний контроль №2	20
Всього підсумкова складова оцінювання:	40
Разом:	100
<p>Розподіл балів за виконання курсової роботи (100 балів):</p> <ul style="list-style-type: none"> • пояснювальна записка – 50 балів; • ілюстративна частина – 10 балів; • захист роботи – 40 балів. <p>Поточне оцінювання та проведення контрольних заходів у межах курсу відбувається згідно нормативних документів НУВГП: Положення про семестровий поточний та підсумковий контроль навчальних досягнень здобувачів вищої освіти https://ep3.nuwm.edu.ua/25889/</p> <p>Положення про атестацію здобувачів вищої освіти та роботу екзаменаційної комісії http://ep3.nuwm.edu.ua/8545/;</p> <p>Порядок ліквідації академічних заборгованостей у НУВГП https://ep3.nuwm.edu.ua/25072/;</p> <p>Положення про навчально-науковий центр незалежного оцінювання Національного університету водного господарства та природокористування http://ep3.nuwm.edu.ua/4184/</p>	
Рекомендована література (основна, допоміжна)	

Рекомендована література

Основна

1. Т.М. Басюк, В.В. Литвин, Л.М. Захарія, Н.Е. Кунанець. *Машинне навчання: Навчальний посібник*. – Львів: Видавництво «Новий Світ - 2000», 2019. - 335 с.
2. *Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних* / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с.
3. *Машинне навчання: методи та моделі: підручник* / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с.
4. *Машинне навчання: комп'ютерний практикум [Електронний ресурс]: навч. посібник* / Л.М. Олещенко – Київ: КПІ ім. Ігоря Сікорського, 2022. – 92 с.
5. *Основи машинного навчання : навч. посіб.* / В. О. Харченко. – Суми : Сумський державний університет, 2023. – 264 с.

Допоміжна

6. *Python data Science Handbook* / Jake VanderPlas. – O'Reilly Media, 2017. – 529 p.
7. *Python for Data Analysis* / Wes McKinney. – O'Reilly Media, 2018. – 522 p.
8. *Python for Probability, Statistics and Machine Learning* / Jose Unpringsco. – Springer, 2016. – 276 p.
9. *Python Machine Learning Cookbook* / Chris Albon. – O'Reilly Media, 2018. – 366 p.

Інформаційні ресурси в Інтернет

1. https://uk.wikipedia.org/wiki/Машинне_навчання
2. <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>
3. https://prometheus.org.ua/course/course-v1:IRF+ML101+2016_T3
4. <http://www.mmf.lnu.edu.ua/ar/1739>
5. <https://www.coursera.org/articles/what-is-machine-learning>
6. <https://github.com/>

Поєднання навчання та досліджень

В процесі навчання здобувачі вищої освіти залучаються до реалізації наукових досліджень у сфері ІТ, зокрема за темою кафедральної НДР «Інформаційні технології моделювання екологічних, економічних та соціальних процесів»; готують доповіді на щорічні університетські та Міжнародні наукові конференції; статті для збірників наукових праць, що видаються в НУВГП (Студентський науковий Вісник, Вісник АКОТ, Вісник НУВГП); беруть участь у студентських олімпіадах, конкурсах студентських наукових робіт, виставках, воркшопах та хакатонах. В навчальному процесі використовуються наукові здобутки лектора, зокрема:

Hrytsiuk P., Babych T., Baranovskii S., Havryliuk M. *Assesing of Climate Impact on Wheat Yield using Machine Learning Techniques. Materials of the XI International Scientific-Practical Conference "Information Control Systems and Technologies" ICST, 2023. pp. 102 – 105.*

ПОЛІТИКИ ВИКЛАДАННЯ ТА НАВЧАННЯ

Перелік соціальних, «м'яких» навичок (soft skills)

Здатність до навчання та саморозвитку
Уміння вчитися впродовж життя
Комунікативність, навички колективної роботи
Комплексне рішення проблем
Критичне мислення
Працелюбність

Дедлайни та перескладання

Крайні терміни захисту лабораторних робіт регламентується останнім тижнем перед початком екзаменаційної сесії. У разі невиконання студентом вимог щодо поточного оцінювання протягом семестру (невчасне виконання) завдання) оцінку може бути знижено в межах 15%.

Ліквідація академічної заборгованості здійснюється згідно «Порядку ліквідації академічних заборгованостей в НУВГП». <https://ep3.nuwm.edu.ua/25072/>. Згідно цього документу і реалізується право студента на повторне вивчення дисципліни чи повторне навчання на курсі.

Оголошення стосовно дедлайнів здачі та перездачі оприлюднюються на сторінці MOODLE

<https://exam.nuwm.edu.ua/course/view.php?id=5947>.

Неформальна та інформальна освіта (за потреби)

Студенти мають право на перезарахування результатів навчання набутих у неформальній та інформальній освіті згідно відповідного положення <https://nuwm.edu.ua/strukturni-pidrozdili/centr-neformalnoji-osviti>.

Студенти можуть самотійно на платформах Prometheus, Coursera, edEx, edEra, Future Learn опанувати матеріал для перезарахування результатів навчання. При цьому важливо, щоб знання та навички, що формуються під час проходження певного онлайн-курсу чи його частин, мали зв'язок з очікуваними програмними результатами навчальної дисципліни та перевірялись в підсумковому оцінюванні.

Перед початком проходження обраних курсів необхідно отримати згоду викладача.

Рекомендовані курси Coursera:

- IBM Машинне навчання з Python: <https://www.coursera.org/learn/machine-learning-with-python>
- Supervised Machine Learning : <https://www.coursera.org/learn/machine-learning>
- Вступ до машинного навчання : <https://www.coursera.org/learn/machine-learning-duke>

Правила академічної доброчесності

У разі виявлення копіювання результатів виконання завдань студенту завдання не зараховується. Студент повторно отримує завдання і виконує його самотійно.

Документи стосовно академічної доброчесності (про плагіат, порядок здачі звіту, кодекс честі студентів, документи Національного агентства стосовно доброчесності) наведені на сторінці НУВГП <http://nuwm.edu.ua/sp/akademichna-dobrochesnistj>

Вимоги до відвідування

Заняття відбуваються згідно розкладу <https://desk.nuwm.edu.ua/cgi-bin/timetable.cgi> офлайн або онлайн за допомогою Google Meet за лінком: <https://meet.google.com/> Консультації проводяться за потреби в режимі онлайн за допомогою Google Meet у домовлений зі студентами час. Здобувачі на заняттях можуть використовувати мобільні телефони та ноутбуки, але виключно в навчальних цілях. Студенту не дозволяється пропускати заняття без поважних причин. За наявності об'єктивних причин пропуску занять, студенти можуть самостійно ознайомитися з теоретичним матеріалом на платформі MOODLE <https://exam.nuwm.edu.ua/course/view.php?id=5947>

Лектор : Грицюк П.М., д.е.н., к.ф.-м.н., професор, завідувач кафедри комп'ютерних технологій та економічної кібернетики

Автор
Завідувач кафедри комп'ютерних
технологій та економічної кібернетики

Петро ГРИЦІЮК

Затверджено

Проректор з науково-педагогічної та
навчальної роботи

Валерій СОРОКА



документ підписаний КЕП
Номер документа СИЛ №868
Підписувач Сорока Валерій Степанович
Підписувач (дані КЕП):
Сертифікат 3FAA9288358EC003040000009B6C3700C8C2C100