

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра системного проектування

Затверджено

На засіданні кафедри СП
Факультету електроніки та
комп'ютерних технологій
Львівського національного
університету імені Івана Франка
(протокол № від серпня 2022 р.)

Силабус з навчальної дисципліни
«Методи та технології машинного навчання»,
що викладається в межах ОПП «Інженерія програмного забезпечення»
першого (бакалаврського) рівня вищої
освіти для здобувачів з спеціальності
121 «Інженерія програмного забезпечення» (ВПК)

Львів 2022

Назва дисципліни	Методи та технології машинного навчання
Адреса викладання дисципліни	м. Львів, вул. Драгоманова, 50
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій, кафедра системного проектування
Галузь знань, шифр та назва спеціальності	121 Інженерія програмного забезпечення (ВПК)
Викладачі дисципліни	Ляшкевич Василь Яремович, доцент
Контактна інформація	vasyl.lyashkevych@lnu.edu.ua , https://electronics.lnu.edu.ua/employee/liashkevych-v-ya
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних занять (за попередньою домовленістю). Також можливі он-лайн консультації через MS Teams або систему електронного навчання Moodle.
Сторінка дисципліни	https://moodle.elct.lnu.edu.ua/course/view.php?id=254
Інформація про дисципліну	Дисципліна «Методи та технології машинного навчання» є нормативною дисципліною з циклу процесійної та практичної підготовки за блоками вибіркових дисциплін для спеціальності 121 Програмна інженерія для освітньої програми «Високопродуктивний комп'ютинг (ВПК)», яка викладається в 6 семестрі в обсязі 3,5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Машинне навчання – це підгалузь систем штучного інтелекту в галузі інформатики. Навчальна програма дисципліни “Методи та технології машинного навчання “визначає зміст і обсяг знань”, необхідних для фахівця в галузі наук про дані та систем штучного інтелекту. Дисципліна охоплює проблематику вивчення сучасного стану технологій машинного навчання, що використовуються для формалізації та обробки даних, вивчення сучасних програмних засобів машинної обробки даних, технологій їх проектування, реалізації та дослідження. У межах навчальної дисципліни “Методи та технології машинного навчання” студенти вивчають методи, що застосовуються для побудови складних моделей та алгоритмів із метою вирішення завдань класифікації, кластеризації та прогнозування.
Мета та цілі дисципліни	Метою вивчення дисципліни є формування у майбутніх фахівців знань та вмінь застосування сучасних методів та засобів розробки, дослідження та використання сучасних технологій аналізу даних на основі машинного навчання для вирішення задач класифікації, регресійного аналізу, прогнозування та ухвалення рішень, а також формування у студентів сучасного наукового світогляду в області

	<p>методів машинного навчання; наукової уяви про задачі, що розв'язуються з допомогою методів машинного навчання, вивчення методів класифікації і регресії з вчителем, а також методи кластерного аналізу (без вчителя); знайомство студентів з сучасними технологіями машинного навчання та тенденціями розробки і застосування; подальше становлення і вдосконалення інформаційної та програмної культури майбутніх фахівців.</p> <p>Основні цілі дисципліни забезпечують: набуття практичних навичок і знань в області технологій машинного навчання. У результаті вивчення даної дисципліни студенти повинні освоїти основні методи навчання з вчителем і без вчителя – кластеризація розбиттям, ієрархічна кластеризація та нечітка кластеризація. Також, в результаті освоєння матеріалу, студенти повинні вивчити основні практичні прийоми роботи з інформацією мовою Python.</p>
<p>Література для вивчення дисципліни</p>	<p>Основна література:</p> <ol style="list-style-type: none"> 1. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn and TensorFlow: O'Reilly, 2017. - 718 p. 2. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MIT Press, 2012. 3. Andrew Ng. Maching Learning Yarning. - [Електронний ресурс]. - Режим доступу: https://nessie.ilab.sztaki.hu/~kornai/2020/AdvancedMachineLearning/Ng_MachineLearningYarning.pdf 4. Robert E. Schapire, Yoav Freund. Boosting: Foundations and Algorithms: MIT Press, 2012. - 544 p. - [Електронний ресурс]. - Режим доступу: https://direct.mit.edu/books/book-pdf/2091763/book_9780262301183.pdf 5. Hastie T., Tibshirani R, Friedman J. The Elements of Statistical Learning (2nd edition). Springer, 2009. 6. Machine learning. A First Course for Engineers and Scientists / Andreas Lindholm, Niklas Wahlström, Fredrik Lindsten, Thomas B. Schön // Cambridge University Press, 2021. - 275 p. 7. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006. 8. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014. 9. Charu C. Aggarwal. Recommender Systems: Springer, 2016. - 518 p. 10. Kishan G. Mehrotra Chilukuri K. Mohan HuaMing Huang. Anomaly Detection Principles and Algorithms: Springer. - 2017. - 229 p. - DOI: https://doi.org/10.1007/978-3-319-67526-8 11. Florentina Hristea, Cornelia Caragea. Natural Language Processing (NLP) and Machine Learning (ML)—Theory and Applications: MDPI, 2022. - 306 p. - DOI: https://doi.org/10.3390/books978-3-0365-5580-5 12. Yue Zhang, Zhiyang Teng. Natural Language Processing: A Machine Learning Perspective, 2021. 13. Machine Learning in Computer Vision / N. Sebe, Ira Cohen, Ashutosh Garg, Thomas S. Huang// Springer, 2005. - 249 p. - [Електронний ресурс]. - Режим доступу: http://silverio.net.br/heitor/disciplinas/eeica/papers/Livros/[Sebe]%-20-%20Machine%20earning%20in%20Computer%20Vision.pdf 14. Cha Zhang, Yunqian Ma. Ensemble Machine Learning: Methods and Applications: Springer, 2012. - 331 p. - DOI: https://doi.org/10.1007/978-1-4419-9326-7
<p>Обсяг курсу</p>	<p>Кількість кредитів ЄКТС: 3,5 (105 год), з них: 64 годин аудиторних занять (лекції: 32 год, лабораторні: 32 год.) та 41 год. самостійної роботи.</p>
<p>Очікувані результати навчання</p>	<p>По завершенні вивчення дисципліни студенти повинні набути наступних компетентностей:</p> <ul style="list-style-type: none"> - ЗК1. Здатність до абстрактного мислення, аналізу та синтезу. - ЗК2. Здатність застосовувати знання у практичних ситуаціях. - ЗК04. Здатність спілкуватися іноземною мовою як усно, так і письмово.

- **ЗК05.** Здатність вчитися і оволодівати сучасними знаннями.
- **ЗК07.** Здатність працювати в команді.
- **ФК14.** Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування.
- **ФК15.** Здатність розробляти архітектури, модулі та компоненти програмних систем.
- **ФК19.** Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.
- **ФК22.** Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя.
- **ФК24.** Здатність здійснювати процес інтеграції системи, застосовувати стандарти і процедури управління змінами для підтримки цілісності, загальної функціональності і надійності програмного забезпечення.
- **ФК25.** Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення.
- **ФК26.** Здатність до алгоритмічного та логічного мислення.
- **ФК27.** Здатність розробляти високопродуктивні програмні комплекси для вирішення завдань наук про дані, систем штучного інтелекту, вбудованих та інших інноваційних систем.
- **ФК28.** Володіння методами розроблення систем підвищеної продуктивності, серверними та розподіленими технологіями, інструментальними засобами проектування та розробки веб-застосунків і нових технологій.
- **ПРН01.** Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.
- **ПРН03.** Знати основні процеси, фази та ітерації життєвого циклу програмного забезпечення.
- **ПРН05.** Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.
- **ПРН06.** Вміння вибирати та використовувати відповідну задачі методологію створення програмного забезпечення.
- **ПРН07.** Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення.
- **ПРН09.** Знати та вміти використовувати методи та засоби збору, формулювання та аналізу вимог до програмного забезпечення.
- **ПРН10.** Проводити передпроектне обстеження предметної області, системний аналіз об'єкта проектування.

	<ul style="list-style-type: none"> - ПРН11. Вибирати вихідні дані для проектування, куруючись формальними методами опису вимог та моделювання. - ПРН13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань. - ПРН15. Мотивовано обирати мови програмування та технології розробки для розв'язання завдань створення і супроводження програмного забезпечення. - ПРН17. Вміти застосовувати методи компонентної розробки програмного забезпечення. - ПРН18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних. - ПРН23. Вміти документувати та презентувати результати розробки програмного забезпечення. - ПРН25. Вміти застосовувати сучасні технологічні рішення щодо розробки програмно-апаратних систем та їх компонентів.
Ключові слова	Машинне навчання, алгоритми машинного навчання, регресія, класифікація, кластеризація, рекомендаційні системи, детектори аномалій, конвеєри машинного навчання.
Формат курсу	Проведення лекцій, лабораторних робіт та консультації для кращого розуміння тем проводиться у формі проектно-орієнтованого підходу з елементами дуальної освіти в компанії ГлобалЛоджик.
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Екзамен в кінці семестру.
Сертифікація	Сертифікація не є обов'язковим елементом дисципліни, а тільки дозволяє оцінити свої можливості для працевлаштування: <ul style="list-style-type: none"> - Python (Basic) Skills - Problem Solving (Basic) Skills - Problem Solving (Intermediate) Skills - Statistics and Machine Learning - Інші навички
Пререквізити	Для вивчення курсу студенти потребують знання та навичок, які можна отримати після проходження дисциплін: “Методи та технології опрацювання даних”, “Методи обчислень” та “Іноземна (англійська) мова”. Знання іноземної вимагається для правильного розуміння і тлумачення термінології в галузі штучного інтелекту і машинного навчання.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентація, лекції, лабораторні роботи, обговорення, дискусія.
Необхідне обладнання	Мультимедійне обладнання, комп'ютерний клас, програми та сервіси MS Teams, Moodle, Python, Numpy, Pandas, Seaborn, Scikit-learn та ін.

<p>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</p>	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою, де враховано бали за два контрольні заміри по 25 балів за кожний модуль та 50 балів за складання екзамену.</p> <p>Бали нараховуються за такими видами робіт з наступним співвідношенням:</p> <ul style="list-style-type: none"> • контрольні заміри (2 модулі): 50% семестрової оцінки; максимальна кількість балів 50, а саме: <ul style="list-style-type: none"> - лабораторні роботи: 64% оцінки контрольного заміру; максимальна кількість балів 32 (16 лабораторних робіт). - теоретичний матеріал: 36% оцінки контрольного заміру; максимальна кількість балів 18 (2 тести по 9 балів кожний). • екзамен: 50% семестрової оцінки; максимально 50 балів. <p>Оцінки за лабораторні заняття розподіляються наступним чином: виконання лабораторних завдань – 50 %, відповіді на запитання викладача або тести – 50 %. В результаті ми маємо 16 лабораторних робіт по 2 бали за кожну, з яких: 1 бал за виконання, а 1 бал за відповідь чи тест.</p> <p>Оцінки за лабораторні заняття розподіляються так:</p> <p>2 (від 0 до 1 бала за виконання, від 0 до 1 бала за тестування/опитування) – студент в повному обсязі володіє навчальним матеріалом, має повне розуміння розглянутої теми, надає правильні відповіді на запитання по темі, код програми функціонує відповідно до завдання;</p> <p>1 (від 0 до 1 бала за виконання, від 0 до 1 бала за тестування/опитування) – студент не досить добре розуміє розглянутий матеріал та написаний ним код програми, вагається та надає неточні/не конкретні відповіді на запитання по темі, код програми функціонує неточно, або з помірними недоліками;</p> <p>0 (0 балів за виконання, 0 балів за тестування/опитування) - студент зовсім не засвоїв розглянутий матеріал, написаний ним код програми не відповідає темі/не функціонує взагалі.</p> <p>Оцінювання екзаменаційних питань:</p> <p>20-25 балів - розглянута тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно, які містять аналіз і систематизацію, аргументовані висновки. Засвідчено глибоке володіння матеріалом. Наведені приклади коду повністю робочі та відповідають темі. Можуть бути присутні несуттєві помилки та невідповідності;</p> <p>15-19 балів - відтворюється значна частина розглянутої теми. Виявлено знання і розуміння основних положень навчальної дисципліни, проте присутні неточності та/або невідповідності основній темі. Наведені приклади коду частково робочі, проте в загальному відповідають темі;</p> <p>10-14 балів - відстежується загальне розуміння розглянутої теми. Виявлені множинні неточності та невідповідності, пояснення наведеного коду відсутні, код функціонує із значними неточностями (або відсутні приклади запуску коду на виконання взагалі);</p> <p>5-9 балів – студент погано розуміє розглянуту тему. Виявлені суттєві неточності та невідповідності. Наведені приклади коду з суттєвими недоліками, або не відповідають темі;</p>
---	---

	<p>0-4 бали – студент взагалі не розуміє розглянуту тему. Тему не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи. Наведений код не робочий, або відсутній як такий.</p> <p>Академічна доброчесність: Очікується, що лабораторні роботи та контрольні роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі студента є підставою для її незарахування викладачем.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції і лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. Студенти зобов'язані дотримуватися усіх термінів визначених для виконання усіх видів робіт дисципліни.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали набрані на поточному тестуванні, самостійній роботі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час лабораторного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін. Жодні форми порушення академічної доброчесності не толеруються.</p>
<p>Питання до контрольних робіт</p>	<p>Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт:</p> <ol style="list-style-type: none"> 1. Поняття даних (<i>Життєвий цикл науки про дані. Різновиди алгоритмів машинного навчання. Типи та види навчання. Типи даних. Структуровані та неструктуровані дані. Категорійні дані. Часові ряди</i>) 2. Основи машинного навчання (<i>Міжгалузевий стандартний процес інтелектуального аналізу даних (CRISP-DM). Основи парадигм навчання. Контрольоване навчання. Навчання без нагляду. Напівконтрольоване навчання. Самоконтрольоване навчання. Навчання з підкріпленням. Важливі групи алгоритмів машинного навчання.</i>) 3. Проект машинного навчання від початку до кінця (<i>Поняття проектного навчання. Ідеологія від початку до кінця. Робота з справжніми даними. Визначення критеріїв для вимірювання продуктивності. Припущення. Гістаграми з даними.</i>) 4. Створення дата-сету (<i>Створення наборів даних для: тренування, доказу та тестування. Пошук кореляції. Візуальний аналіз даних. Типи алгоритмів машинного навчання. Поняття моніторингу та підтримки системи.</i>)

5. Основи статистичного аналізу даних (*Середнє арифметичне, Медіана, Мода, Середньоквадратичне відхилення, Дисперсія, Математичне очікування, Розподіли даних. Гістограми.*)
6. Основи теорії ймовірностей (*Визначення ймовірності. Статистичний підхід. Поняття випадковості в теорії ймовірності та статистиці. Операції над подіями. Теорема додавання. Незалежність подій. Умовна ймовірність. Теорема Байєса.*)
7. Використання теорії ймовірностей (*Випадкові величини. Функція розподілу. Дискретні випадкові величини. Зв'язок між розподілами. Безперервні величини. Нормальний розподіл. Багатомірний нормальний розподіл*)
8. Вимоги до проекту машинного навчання (*Поняття проекту машинного навчання у промисловості. Вимоги до оточення інфраструктури машинного навчання. Категоризація проблем з даними. Крайні практики для різних проблем даних. Анотування даних. Життєвий цикл проекту з машинним навчанням. Приклад планування проекту для розпізнавання мови.*)
9. Доставка моделей машинного навчання (*Поняття доставки: підходи, ключові задачі. Концепція зсуву та зсув даних. Типові шляхи розгортання систем з машинними навчанням. Система моніторингу. Система керування моніторингом. Метрики та їх ознаки. Ітераційна доставка. Конвеєр моніторингу. Важливість даних. Конвеєри даних. Збір та моніторинг даних. Підготовка дата-сетів. Поняття крос-валідації. Типові проблеми з даними.*)
10. Ймовірнісний підхід до лінійних моделей (*Ймовірнісний підхід. Проблема класифікації. Функції втрат. Класифікація та логістична регресія. Описові та генеративні алгоритми. Узагальнена лінійна модель (GLM). Типи ймовірностей. Генеративні моделі. Типи ймовірностей. Умовні ймовірності. Теорема Байєса. Класифікатор Байєса.*)
11. Дерева рішень (*Огляд дерев рішень. Поняття ентропії. Навчання дерев рішень. Проблема перенавчання. Застосування дерев рішень. Переваги та недоліки дерев рішень.*)
12. Поняття задач регресії (*Огляд регресії. Різниця між моделлю регресії та математичною функцією. Побудова регресійної моделі. Поліноміальні моделі регресії. Bias-variance tradeoff. Незалежний вибір змінної. Формулювання регресійної моделі. Оцінка. Перевірка. Прогноз / Застосування.*)
13. Особливості візуалізації даних (*Декомпозиція та агрегація даних. Трансформація даних. Поділ даних. Злиття даних. Групування. Сортування. Matplotlib. Фігури. Полотно. Візуалізація.*)
14. Візуалізація засобами Python (*Налаштування фігури. Pandas. Pandas об'єкти. Bar plot. Stacked Bar Plots. Seaborn. Агрегація та сумаризація. Гістограми та розподіл густини. Розкиди або точкові діаграми. Фасетні сітки та категорійні дані.*)
15. Поняття інженерії ознак (*Цілі і задачі інженерії ознак. Базові типи ознак. Ресурси даних. Загальні підходи до отримання ознак тексту. Токенізація. Bag of Words. Нормалізація. Вибір 6 ознак. Патерни інженерії ознак. Ітерація та вдосконалення (робота з помилками).*)
16. Аналіз помилок в процесі навчання (*Аналіз помилок. Поняття дискримінаторів ознак. Теорія рішень. Використання теореми Байєса. Використання статистичних методів для видобування ознак. Особливості видобування ознак з малюнків.*)

17. Дослідження ознак даних (*Поняття оглядового аналізу даних. Засоби для аналізу даних. Інженерія ознак. Поняття ознаки. Трансформація ознак. Масштабування даних.*)
18. Дослідження важливості ознак даних із Scikit-learn (*Поняття ознак для структурованих та не структурованих даних. Важливість ознак. Методи визначення важливостей ознак. Використання Scikit-learn для визначення важливості ознак.*)
19. Колеляція в ознаках даних (*Поняття коваріації. Матриця коваріації. Змінні, що корелюють. Лінійна кореляція. Кореляція Пірсона. Вивчення лінійних залежностей. Метод спроб і помилок. Оптимізація. Кореляція за багатьма змінними. Залишки регресії. Матриця кореляції. Поняття коваріації*)
20. Поняття регуляризації (*Поліноміальна регресія. Пошук проблеми моделі чи даних. Зважена лінійна регресія. Узагальнені моделі. Компроміс між зміщенням та варіацією (Bias-variance trade-off). Регуляризація. L1, L2-регуляризація. Особливості регресії для вибору ознак.*)
21. Оптимізація у машинному навчанні (*Математична оптимізація. Похідна. Критичні точки. Градієнтний спуск. Метрики Якобі та Хесіяна. Метод Ньютона. Градієнтний спуск без градієнту. Обмежена оптимізація. Інші методи для застосування.*)
22. Поняття метрик якості (*Поняття оціночної метрики. Метрики оцінки вирішення задач регресії, класифікації, навчання без учителя, на ін. Метрика точності, акуратності, f1-score.*)
23. Оцінки та методи розрахунку метрик (*Розрахунок метрик. Задовільнення критеріїв та оптимізація метрик. Упередженість, якої можна уникнути. Типові помилки даних. Поняття аналізу помилок. Інтерпретація помилки. Інтерпретаційні моделі. Властивості інтерпретаційних моделей.*)
24. Лінійні моделі (*Машинне навчання. Базові визначення. Контрольоване навчання. Навчання без контролю. Набір для навчання, набір для перевірки та набір для тестування. Переобладнання та недообладнання. Лінійні моделі. Градієнтний спуск. Стохастичний градієнтний спуск.*)
25. Машина опорних векторів (*Поняття машини опорних векторів (SVM). Лінійно роздільні зразки. Нелінійно роздільні зразки. Класифікація опорних векторів. Класифікація м'якої маржі. Нелінійна класифікація SVM. Поліноміальне ядро. Додавання ознак подібності. Гауссове ядро RBF. Підтримка векторної регресії. Функція прийняття рішень і прогнози. Квадратичне програмування. Ядро SVM.*)
26. Поняття Boosting (*Проблеми класифікації у машинному навчанні. Бустінг. Опір до перенавчання. Теорія границь. Основні алгоритми. Загальний алгоритм аналізу помилок. AdaBoost. Мінімізація помилок. Проектування базових алгоритмів навчання. Ефективність бустингу. Бустінг для часових рядів.*)
27. Поняття нейронної мережі (*Логістична регресія. Функція витрат. Перцептрон Розенблата. XOR проблема. Багатошарові мережі. Теорема універсальної апроксимації. Поняття обчислювального графу. Правила ланцюжка. Похідна. Обчислювальних граф логістичної регресії. Пряме і зворотнє розповсюдження помилки. Векторизація. Дерева та ліси. Глибинні ознаки. Поняття нейронної мережі на основі перцептрона. Багатошарові перцептрони.*)
28. Навчання без учителя (*Поняття навчання без учителя. Кластеризація. Скорочення розрядності. Приклади застосування*

	<p><i>кластеризації. K-середніх. Модель суміші Гауса. Ієрархічна кластеризація. Методи на основі щільності. Спектральна кластеризація. Проблеми високої розрядності.)</i></p> <p>29. <i>Поняття системи рекомендацій (Що таке система рекомендацій? Навіщо потрібні системи рекомендацій? Завантаження даних. Роль дослідницького аналізу даних. Що таке рейтинги? Середній Байєс. Трансформація даних. Обчислення розрідженості матриці.)</i></p> <p>30. <i>Побудова системи рекомендацій (Загальні поняття. Огляд систем рекомендацій. Побудова систем рекомендацій на основі методів машинного навчання. Генерація кандидатів. Міри та моделі 7 схожості. Принципи аналізу вхідних даних. Доменно-специфічні виклики в системах рекомендацій. Використання фільтрів. Матриця фактуризації. Масштабний пошук. Побудова гіпотез для заповнення пропусків у даних. Побудова конвеєру машинного навчання на основі табличних даних.)</i></p> <p>31. <i>Поняття та пошук аномалій (Поняття аномалій. Аномалії в індустрії. Методи і алгоритми пошуку аномалій. Оцінка густини. Підходи до виявлення аномалій на основі відстані. Пошук схожих об'єктів. Міри схожості. Спільна фільтрація. Фільтрування на основі вмісту.)</i></p> <p>32. <i>Системи виявлення аномалій (Підходи до виявлення аномалій на основі кластеризації. Підходи до виявлення аномалій на основі моделі. Алгоритми на основі відстані та щільності. Рангові алгоритми.)</i></p> <p>33. <i>Машинне навчання для NLP (Поняття опрацювання природньої мови. Нормалізація тексту. Векторне представлення слів. Word2Vec і Word Embeddings. Вирішення питань опрацювання текстових даних. Поняття векторного простору. Ознаки текстів. Класифікація та кластеризація текстів. Організація конвеєрів для опрацювання текстів.)</i></p> <p>34. <i>Машинне навчання та комп'ютерний зір (Малюнок для машинного навчання. Принципи опрацювання зображень. Поняття ознак зображень. Екстрактори ознак. Розпізнавання об'єктів та зображень. Класифікація та кластеризація зображень.)</i></p> <p>35. <i>Засоби машинного навчання для комп'ютерного зору (Сегментація та семантична сегментація. Вирішення питань комп'ютерного зору в індустрії. Засоби для роботи з малюнками. Побудова конвеєру для опрацювання зображень.)</i></p> <p>36. <i>Поняття ансамблю у машинному навчанні (Навчання в ансамблі. Алгоритми підвищення: огляд методів, теорії та застосунки. Підвищення оцінки ядра. Цільове навчання. Випадкові ліси. Навчання ансамблю за допомогою навчання негативної кореляції. Дискримінаційне навчання для анатомічної структури. Виявлення та сегментація.)</i></p>
<p>Опитування</p>	<p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p>

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності	Л-ра. Ресурси в Інтернеті.	Завдання, год	Термін виконання, тиж.
1	Основи машинного навчання. Міжгалузевий стандартний процес інтелектуального аналізу даних (CRISP-DM). Основи парадигм навчання. Контрольоване навчання. Навчання без нагляду. Напівконтрольоване навчання. Самоконтрольоване навчання. Навчання з підкріпленням. Важливі групи алгоритмів машинного навчання.	лекція	1-3	2	кінець поточного тижня
	Лаб 1. Знайомство із бібліотекою scikit-learn	лаб. робота	1	2	кінець поточного тижня
	Застосування машинного навчання в індустрії	сам. робота	1	2,56	кінець поточного тижня
2	Поняття проекту машинного навчання. Поняття проектного навчання. Ідеологія від початку до кінця. Робота з справжніми даними. Визначення критеріїв для вимірювання продуктивності. Припущення. Гістаграми з даними. Створення наборів даних для: тренування, доказу та тестування. Пошук кореляції. Візуальний аналіз даних. Типи алгоритмів машинного навчання. Поняття моніторингу та підтримки системи.	лекція	1, 3	2	кінець поточного тижня
	Лаб 2. Підготовка та опрацювання табличних дата-сетів	лаб. робота	1, 2	2	кінець поточного тижня
	Поняття та вирішення задач регресії	сам. робота	1	2,56	кінець поточного тижня
3	Проектування рішення задачі на основі машинного навчання. Вимоги до оточення інфраструктури машинного навчання. Поняття проекту машинного навчання у промисловості. Категоризація проблем з даними. Анотування даних. Життєвий цикл проекту з машинним навчанням. Приклад планування проекту для розпізнавання мови. Поняття доставки: підходи, ключові задачі. Концепція зсуву та зсув даних. Типові шляхи розгортання систем з машинними навчанням. Система керування моніторингом. Метрики та їх ознаки. Ітераційна	лекція	1-3	2	кінець поточного тижня

	доставка. Конвеєр моніторингу. Важливість даних. Конвеєри даних. Збір та моніторинг даних. Підготовка дата-сетів. Поняття крос-валідації. Типові проблеми з даними.				
	Лаб 3. Вирішення завдань лінійної регресії	лаб. робота	1-3	2	кінець поточного тижня
	Поняття та вирішення задач класифікації	сам. робота	1-4	2,56	кінець поточного тижня
4	Навчання імовірнісних моделей. Ймовірнісний підхід до лінійних моделей. Ймовірнісний підхід. Проблема класифікації. Функції втрат. Класифікація та логістична регресія. Описові та генеративні алгоритми. Узагальнена лінійна модель (GLZ). Типи імовірностей. Генеративні моделі. Типи імовірностей. Умовні імовірності. Теорема Байєса. Класифікатор Байєса. Модель Маркова. Прихована модель Маркова.	лекція	5-6	2	кінець поточного тижня
	Лаб 4. Опрацювання вибраних елементів теорії ймовірності	лаб. робота	5	2	кінець поточного тижня
	Поняття та вирішення задач кластеризації	сам. робота	1-3	2,56	кінець поточного тижня
5	Дерева рішень та регресія. Огляд дерев рішень. Поняття ентропії. Навчання дерев рішень. Проблема перенавчання. Застосування дерев рішень. Переваги та недоліки дерев рішень. Огляд регресії. Різниця між моделлю регресії та математичною функцією. Побудова регресійної моделі. Поліноміальні моделі регресії. Bias-variance tradeoff. Незалежний вибір змінної. Формулювання регресійної моделі. Оцінка. Перевірка. Прогноз / Застосування.	лекція	2, 6	2	кінець поточного тижня
	Лаб 5. Дерева рішень та регресія	лаб. робота	5-6	2	кінець поточного тижня
	Поняття та впровадження проекту машинного навчання	сам. робота	2, 3	2,56	кінець поточного тижня
6	Пошук оптимального набору ознак. Цілі і задачі інженерії ознак. Базові типи ознак. Ресурси даних. Загальні підходи до отримання ознак тексту. Токенізація. Bag of Words. Нормалізація. Вибір ознак. Патерни інженерії	лекція	1, 6	2	кінець поточного тижня

	<p>ознак. Ітерація та вдосконалення (робота з помилками). Аналіз помилок. Поняття дискримінаторів ознак. Теорія рішень. Використання теореми Байеса. Використання статистичних методів для видобування ознак. Особливості видобування ознак з малюнків.</p>				
	Лаб 6. Вибір найкращих ознак для даних	лаб. робота	1	2	кінець поточного тижня
	Методи вирішення задач машинного навчання	сам. робота	1, 3	2,56	кінець поточного тижня
7	<p>Регуляризація та оптимізація. Поліноміальна регресія. Пошук проблеми моделі чи даних. Зважена лінійна регресія. Узагальнені моделі. Копроміс між зміщенням та варіацією (Bias-variance trade-off). Регуляризація. L1, L2-регуляризація. Особливості регресії для вибору ознак. Математична оптимізація. Похідна. Критичні точки. Градієнтний спуск. Метрики Якобі та Хесіяна. Метод Ньютона. Градієнтний спуск без градієнту. Обмежена оптимізація. Інші методи для застосування.</p>	лекція	1-3	2	кінець поточного тижня
	Лаб 7. Оптимізація лінійних моделей градієнтним спуском	лаб. робота	1-3	2	кінець поточного тижня
	Технології вирішення задач машинного навчання	сам. робота	1-3	2,56	кінець поточного тижня
8	<p>Навчання лінійних моделей. Лінійні моделі. Машинне навчання. Базові визначення. Контрольоване навчання. Навчання без контролю. Набір для навчання, набір для перевірки та набір для тестування. Переобладнання та недообладнання. Лінійні моделі. Градієнтний спуск. Стохастичний градієнтний спуск.</p>	лекція	1, 3	2	кінець поточного тижня
	Лаб 8. Побудова власних моделей та їх оцінювання	лаб. робота	2, 8	2	кінець поточного тижня
	Маштабування задач машинного навчання	сам. робота	2, 3, 8	2,58	кінець поточного тижня
9	<p>Поняття Boosting. Проблеми класифікації у машинному навчанні. Бустінг. Опір до перенавчання. Теорія границь. Основні алгоритми. Загальний алгоритм аналізу помилок. AdaBoost.</p>	лекція	4	2	кінець поточного тижня

	Мінімізація помилок. Проектування базових алгоритмів навчання. Ефективність бустингу. Бустінг для часових рядів.				
	Лаб 9. Дослідження алгоритму AdaBoost	лаб. робота	1, 3	2	кінець поточного тижня
	Поняття ядр та функціональних просторів	сам. робота	1, 3	2,56	кінець поточного тижня
10	Передумови нейронних мереж. Логістична регресія. Функція витрат. Перцептрон Розенблата. XOR проблема. Багатошарові мережі. Теорема універсальної апроксимації. Поняття обчислювального графу. Правила ланцюжка. Похідна. Обчислювальних граф логістичної регресії. Пряме і зворотне розповсюдження помилки. Векторизація. Дерева та ліси. Глибинні ознаки. Поняття нейронної мережі на основі перцептрона. Багатошарові перцептрони.	лекція	1, 3, 4	2	кінець поточного тижня
	Лаб 10. Дослідження алгоритмів класифікації	лаб. робота	1, 3	2	кінець поточного тижня
	Побудова конвеєрів машинного навчання на основі великих даних	сам. робота	1, 3, 6-8	2,56	кінець поточного тижня
11	Навчання без учителя. Поняття навчання без учителя. Кластеризація. Скорочення розрядності. Приклади застосування кластеризації. К-середніх. Модель суміші Гауса. Ієрархічна кластеризація. Методи на основі щільності. Спектральна кластеризація. Проблеми високої розрядності. Методи скорочення розрядності.	лекція	1, 6	2	кінець поточного тижня
	Лаб 11. Дослідження алгоритмів кластеризації	лаб. робота	1, 6-8	2	кінець поточного тижня
	Алгоритми машинного навчання	сам. робота	1, 6-8	2,56	кінець поточного тижня
12	Системи рекомендацій. Загальні поняття. Огляд систем рекомендацій. Побудова систем рекомендацій на основі методів машинного навчання. Генерація кандидатів. Міри та моделі схожості. Принципи аналізу вхідних даних. Доменно-специфічні виклики в системах рекомендацій. Використання фільтрів. Матриця фактуризації. Масштабний пошук. Побудова гіпотез для	лекція	9	2	кінець поточного тижня

	заповнення пропусків у даних. Побудова конвеєру машинного навчання на основі табличних даних.				
	Лаб 12. Побудова системи рекомендацій	лаб. робота	9	2	кінець поточного тижня
	Алгоритми машинного навчання з Scikit-learn	сам. робота	1, 6, 9-10	2,56	кінець поточного тижня
13	Проектування детектору аномалій. Поняття аномалій. Аномалії в індустрії. Методи і алгоритми пошуку аномалій. Оцінка густини. Підходи до виявлення аномалій на основі відстані. Підходи до виявлення аномалій на основі кластеризації. Підходи до виявлення аномалій на основі моделі. Алгоритми на основі відстані та щільності. Рангові алгоритми.	лекція	10	2	кінець поточного тижня
	Лаб 13. Дослідження алгоритмів пошуку аномалій	лаб. робота	10	2	кінець поточного тижня
	Дослідження важливості ознак	сам. робота	1, 10	2,56	кінець поточного тижня
14	Машинне навчання для NLP. Поняття опрацювання природньої мови. Нормалізація тексту. Векторне представлення слів. Word2Vec і Word Embeddings. Вирішення питань опрацювання текстових даних. Поняття векторного простору. Ознаки текстів. Класифікація та кластеризація текстів. Організація конвеєрів для опрацювання текстів.	лекція	1, 3, 8	2	кінець поточного тижня
	Лаб 14. Побудова класифікатора спаму	лаб. робота	1, 3, 8	2	кінець поточного тижня
	Побудова конвеєрів машинного навчання	сам. робота	1, 3, 8	2,56	кінець поточного тижня
15	Машинне навчання та комп'ютерний зір. Малюнок для машинного навчання. Принципи опрацювання зображень. Поняття ознак зображень. Екстрактори ознак. Розпізнавання об'єктів та зображень. Класифікація та кластеризація зображень. Відновлення якості. Сегментація та семантична сегментація. Вирішення питань комп'ютерного зору в індустрії. Засоби для роботи з малюнками. Побудова конвеєру для опрацювання зображень.	лекція	1, 3, 11	2	кінець поточного тижня

	Лаб 15. Класифікація графічних зображень	лаб. робота	11	2	кінець поточного тижня
	Побудова конвеєрів машинного навчання систем кластеризації даних	сам. робота	1, 3, 8, 11	2,56	кінець поточного тижня
16	Поняття ансамблю у машинному навчанні. Навчання в ансамблі. Алгоритми підвищення: огляд методів, теорії та застосунки. Підвищення оцінки ядра. Цільове навчання. Випадкові ліси. Навчання ансамблю за допомогою навчання негативної кореляції. Дискримінаційне навчання для анатомічної структури. Виявлення та сегментація.	лекція	1, 4, 14	2	кінець поточного тижня
	Лаб 16. Масштабування моделей машинного навчання	лаб. робота	1, 3, 8	2	кінець поточного тижня
	Використання ансамблів	сам. робота	1, 3, 4, 14	2,56	кінець поточного тижня