

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра радіоелектронних і комп'ютерних систем

Затверджено

На засіданні кафедри РКС
факультету електроніки та комп'ютерних
технологій
Львівського національного університету
імені Івана Франка
(протокол № ____ від 31 серпня 2020 р.)

В.о. завідувача кафедри

Силабус з навчальної дисципліни
«Машинне навчання»,
що викладається в межах ОПП «Інформаційні системи та
технології» першого (бакалаврського) рівня вищої освіти для
здобувачів з спеціальності
126 «Інформаційні системи та технології»

Назва дисципліни	Машинне навчання
Адреса викладання дисципліни	м. Львів, вул. Драгоманова, 50
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій, кафедра радіоелектронних і комп'ютерних систем
Галузь знань, шифр та назва спеціальності	12 Інформаційні технології, 126 Інформаційні системи та технології
Викладачі дисципліни	Сінькевич Олег Олександрович, асистент
Контактна інформація викладачів	oleh.sinkevych@lnu.edu.ua, https://electronics.lnu.edu.ua/employee/o_sinkevych
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних занять (за попередньою домовленістю). Також можливі он-лайн консультації через MS Teams та Telegram (у текстовому режимі). Для погодження часу онлайн консультацій слід писати на електронну пошту викладача, або в особистий Telegram.
Сторінка дисципліни	https://electronics.lnu.edu.ua/course/mashynne-navchannia-ist
Інформація про дисципліну	Дисципліна «Машинне навчання» є вибірковою дисципліною з спеціальності 126 Інформаційні системи та технології для освітньої програми «Інформаційні системи та технології», яка викладається в 6 семестрі в обсязі 6,0 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено таким чином, щоб надати студентам необхідні знання на практичні навички, обов'язкові для того, щоб вміти реалізовувати та застосовувати типові алгоритми та відповідні їм структури даних для вирішення задач машинного навчання. Тому у дисципліні представлено як теоретичні та математичні основи моделей машинного навчання (лінійна та логістична регресія, нейронні мережі, метод головних компонент, k-середніми та ієрархічної кластеризації), так і конкретні реалізації та бібліотеки, які необхідні для створення прототипів та прикладних програмних інтерфейсів (API). Зокрема, детально розглянуті кроки для інженерії ознак, визначення ефективності моделей та підготовки моделей до їх веб-розгортання.
Мета та цілі дисципліни	Метою вивчення вибіркової дисципліни «Мова програмування Python» є одержання студентами знань про базові моделі машинного навчання; типи машинного навчання; процес підготовки та обробки даних; виділення значущих ознак даних, їх перетворення; процес розгортання моделей машинного навчання та їх огортання в прикладний програмний інтерфейс; ознайомлення з бібліотеками машинного навчання для мови Python 3: numpy, scikit-learn, pandas.
Література для вивчення дисципліни	Основна література: <ol style="list-style-type: none"> 1. Эндрю Траск. Грокаем глубокое обучение. СПб.: Питер, 2019. — 352 с. 2. Andreas C. Müller, Sarah Guido. Introduction to Machine Learning with Python. O'Reilly Media, 2017. — 377 с. 3. М. Гринберг. Разработка веб-приложений с использованием Flask на языке Python. ДМК Пресс. - 272 с. 4. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer. — 2009. — 809 с. 5. Peter Harrington. Machine Learning in Action. Manning Publications Co. — 2012. — 382 с.

	<p>6. Luis Pedro Coelho, Willi Richert. Building Machine Learning Systems with Python, 2 edition. Packt Publishing. — 2015. — 326 с.</p> <p>7. Sebastian Raschka, Vahid Mirjalili. Python Machine Learning, 2 edition. Packt Publishing. — 2017. — 622 с.</p> <p>8. Rui Xu, Don Wunsch. Clustering. Wiley-IEEE Press: IEEE Press Series on Computational Intelligence. — 2008. — 364 с.</p> <p>9. Pramod Singh. Deploy Machine Learning Models to Production: With Flask, Streamlit, Docker, and Kubernetes on Google Cloud Platform. Apress. — 2021. — 161 с.</p> <p>10. Machine Learning Mastery [Електронний ресурс] – Режим доступу до ресурсу: https://machinelearningmastery.com/.</p> <p>11. Jones A. The Unsupervised Learning Workshop / A. Jones, K. Christopher, B. Johnston., 2020. – 549 с. – (Packt Publishing).</p> <p>12. Rokach L. Data mining with decision trees. Theory and Applications / L. Rokach, O. Maimon. – Singapore: World Scientific Publishing Co. Pte. Ltd., 2015. – 328 с.</p>
Обсяг курсу	64 години аудиторних занять. З них 32 години лекцій, 32 години лабораторних робіт та 90 годин самостійної роботи.
Очікувані результати навчання	<p>Після завершення цього курсу студент буде:</p> <ul style="list-style-type: none"> - Знати процес створення “поточної лінії” машинного навчання - від обробки даних до розгортання моделі; основи обробки даних, їх перетворення та підготовки для моделей машинного навчання; типові алгоритми виявлення та оцінки ознак даних; основи навчання моделей та оцінки їх ефективності; типові програмні засоби машинного навчання; теоретичний та практичний матеріал згідно програми курсу: математичні основи моделей машинного навчання, програмні реалізації тощо. - Вміти застосовувати засвоєний матеріал для створення прототипів моделей машинного навчання; використовувати Python для програмної реалізації моделей машинного навчання, їх тестування та оцінки, а також визначати тип задачі машинного навчання: регресії, класифікації та кластеризації і розв’язувати поставлену проблему на основі здобутих знань та навиків.
Ключові слова	Python, машинне навчання, логістична регресія, кластеризація, класифікація, нейронні мережі.
Формат курсу	Очний
	Проведення лекцій, лабораторних робіт та консультації для кращого розуміння тем.
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Залік в кінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань з дисциплін «Алгоритми та структури даних», «Основи програмування», «Дискретна математика».
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентація, лекції, лабораторні роботи, обговорення, дискусія.
Необхідне обладнання	Мультимедіа, платформа Moodle, комп’ютерне програмне забезпечення.
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою. Бали нараховуються за такими видами робіт з наступним співвідношенням:</p> <ul style="list-style-type: none"> • лабораторні роботи: 60% семестрової оцінки; максимальна кількість балів 60.

	<ul style="list-style-type: none"> • контрольні заміри (2 модулі): 30% семестрової оцінки; максимальна кількість балів 30. • відвідування занять: 10% семестрової оцінки; максимальна кількість балів 10. <p>Загалом упродовж семестру 100 балів.</p> <hr/> <p>Контрольні заміри проводяться у формі тестових завдань. Академічна доброчесність: Очікується, що лабораторні та контрольні роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикавання 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, незалежно від масштабів плагіату чи обману.</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, передбачених курсом.</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 та самостійній роботі. 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.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
Питання до контрольних робіт	Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт розміщені на веб-сторінці.
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література. Ресурси в Інтернеті	Завдання (лабораторна робота), год	Термін виконання
1	Основи машинного навчання. Типи машинного навчання. Інтелектуальний аналіз даних. Основні поняття. Категорії машинного навчання. Навчання з учителем, без учителя та навчання з підкріпленням. Виявлення патернів у даних та їх використання у розробці алгоритмів прогнозування. Pipeline проєктів машинного навчання.	Лекція	2, 4, 5, 10		
2	Параметричні та непараметричні моделі. Алгоритм k-найближчих	Лекція	2, 4, 6, 10	Реалізація алгоритму k-найближчих сусідів	3 тиж. семестру

	<p>сусідів. Поняття класифікації та регресії. Параметричні та “жадібні” моделі. Поняття навчання моделі з точки зору її параметрів. Непараметричний метод k-найближчих сусідів для задачі класифікації. Переваги та недоліки. Поняття регресії проти класифікації.</p>			з використанням numpy та pandas, 3 год.	
3	<p>Лінійна регресія. Основи лінійної алгебри для векторизації та роботи з матрицями в алгоритмах машинного навчання. Метод найменших квадратів. Функція втрат. Виведення рівняння лінійної регресії. Робота з бібліотекою Numpy для матричних і векторних обчислень. Векторизація задачі лінійної регресії. Визначення методу найменших квадратів та його застосування у машинному навчанні. Поняття функції втрат.</p>	Лекція	4, 10	Створення лінійно-регресійної моделі для багатьох ознак з використанням пакету numpy, 4 год.	4 тиж. семестру
4	<p>Логістична регресія (ЛР). Представлення гіпотези. Мультикласова класифікація. Ймовірнісна інтерпретація ЛР. Задача класифікації методом лінійної регресії. Сигмоїд. Представлення задачі з використанням Numpy. Ймовірнісна інтерпретація логістичної регресії. Алгоритм реалізації. Зведення до задачі з багатозначною цільовою змінною.</p>	Лекція	2, 4, 5, 6, 10	Реалізація логістичної регресії, 2 год.	5 тиж. семестру
5	<p>Аналіз ефективності моделей ЛР. Навчальна крива. Основи регуляризації. Оцінка ефективності навчання моделі логістичної регресії. Візуалізація навчальної кривої. Поняття перенавчання моделей. Перехресне затвердження та k-кратне перехресне затвердження. Регуляризація моделі.</p>	Лекція	2, 4, 5, 7, 10	Модифікація логістичної регресії регуляризаційним доданком. Вивчення кривої навчання, 3 год.	6 тиж. семестру
6	<p>Інженерія ознак. Вибір ознак та аналіз їх значущості. Метод головних компонент. Вибір та інженерії ознак даних у машинному навчанні. Критерії значущості ознак. Алгоритми вибору ознак для задач машинного навчання. Стискання вимірності та метод головних компонент.</p>	Лекція	2, 3, 10	Відбір та виявлення важливості ознак. Метод головних компонент, 4 год.	7 тиж. семестру
7	<p>Нейронні мережі (НМ). Математична основа НМ. Матричне представлення. Функції активація та втрат. Визначення штучної нейронної мережі. Матричне представлення</p>	Лекція	1, 7		

	з використанням NumPy. Поняття функції активації, їх типи. Вибір функції втрат: бінарна та багато класова класифікація. Різновиди сучасних нейронних мереж та їх застосування.				
8	Навчання нейронних мереж. Метод зворотного поширення помилки. Особливості ініціалізації ваг. Навчання нейронних мереж: алгоритм градієнтного спуску та його модифікації. Покроковий алгоритм методу зворотного поширення помилки. Початкове налаштування гіпер-параметрів та ініціалізація ваг. Функція активації у процесі навчання	Лекція	1, 7, 10	Реалізація штучної нейронної мережі з використанням numpy. Навчання методом зворотного поширення помилки, 4 год.	9 тиж. семестру
9	Основи глибокого навчання. Рекурентні (CNN), LSTM, Трансформери. Проблема оптимізації гіпер-параметрів. Багатошарові нейронні мережі та глибоке навчання. Поняття рекурентної нейронної мережі та довгої короткочасної пам'яті. Трансформери та масштабована скалярно-добуткова увага. Гіпер-параметри, проблеми та алгоритми їх оптимізація.	Лекція	1, 6, 7, 10		
10	Дерева прийняття рішень. Індекс gini. Побудова класифікаційних і регресійних дерев. Метод випадкового лісу. Алгоритм дерева прийняття рішень. Поняття індексу gini. Алгоритм CART. Ітераційна побудова дерева для задач класифікації та регресії. Перенавчання дерева прийняття рішень та ансамблевий метод випадкового лісу. Огляд алгоритму випадкового лісу.	Лекція	10, 12		
11	Основи підсилювання. Мета-алгоритми. Adaboost та xgboost методи. Поняття градієнтного підсилення та мета-алгоритмів. Слабкі класифікатори та їх навчання. Алгоритми adaboost. Бутстреп-агрегування (бегінг). Бібліотека xgboost та його використання у Python-інфраструктурі машинного навчання.	Лекція	10, 12	Дослідження дерев прийняття рішень. Порівняння з методами випадкового лісу та градієнтного підсилення, 4 год.	13 тиж. семестру
12	Навчання без учителя. Кластеризація. Метод k-середніх. Оцінка моделі та	Лекція	8, 10, 11	Реалізація кластеризації методом k-середніх.	15 тиж. семестру

	<p>зниження розмірності даних. Основи навчання без учителя та його види: кластеризація та визначення асоціацій. Алгоритм методи k-середніх: підготовка даних (нормалізація/стандартизація), реалізація алгоритму та оцінка оптимальності. Застосування методів головних компонент та t-SNE.</p>			Визначення оптимального параметру k, 4 год.	
13	<p>Ієрархічна кластеризація. Інтеграція моделей з MongoDB. MongoDB pipeline. Ієрархічна кластеризація. Побудова дендрограми та оцінка кількості кластерів. Типи зчеплення (linkage). Агломеративна та розділююча кластеризація. Переваги та недоліки. Збереження даних у NoSQL базі даних MongoDB.</p>	Лекція	8, 10, 11		
14, 15	<p>Основи розгортання моделей. Flask і Fastapi. Огортання моделей машинного навчання в API. Огортання моделей машинного навчання у Flask/Fastapi. Створення Restapi для моделей машинного навчання. Структурування коду. Реалізація простого full pipeline засобами scikit-learn, MongoDB та Flask/Fastapi.</p>	Лекція	3, 9, 10	Розгортання моделі з використанням мікро-фреймворків Flask, або Fastapi API, 4 год.	16 тиж. семестру
16	<p>Підсумкове заняття. Підведення підсумків. Наставочі питання.</p>	Лекція	1, 9, 10		